Signal Analysis Using Autoregressive Models of Amplitude Modulation

Sriram Ganapathy

Advisor - Hynek Hermansky

Johns Hopkins University

11-18-2011

Overview

- Introduction
- AR Model of Hilbert Envelopes
- FDLP and its Properties
- Applications
- Summary

Overview

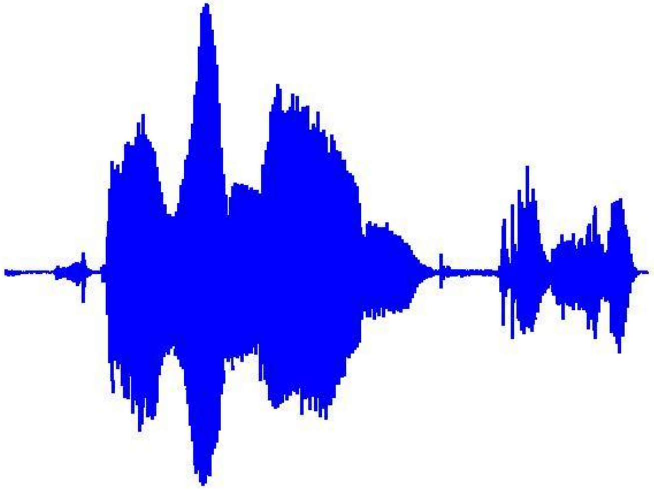
- Introduction
- AR Model of Hilbert Envelopes
- FDLP and its Properties
- Applications
- Summary

Introduction

- Sub-band speech and audio signals - product of smooth modulation with a fine carrier.

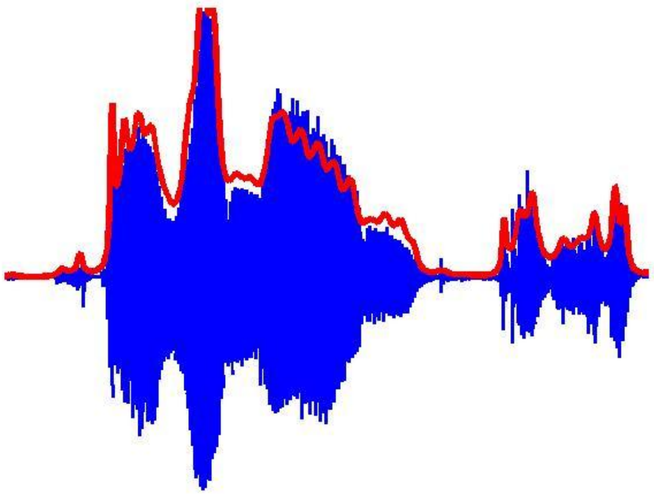
Introduction

- Sub-band speech and audio signals - product of smooth modulation with a fine carrier.



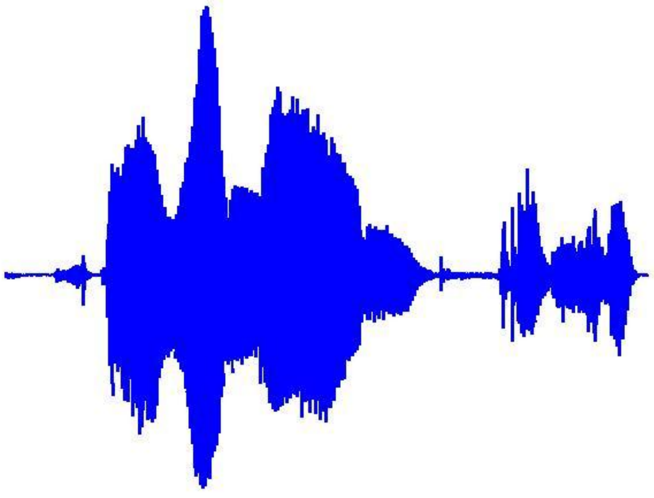
Introduction

- Sub-band speech and audio signals - product of smooth modulation with a fine carrier.



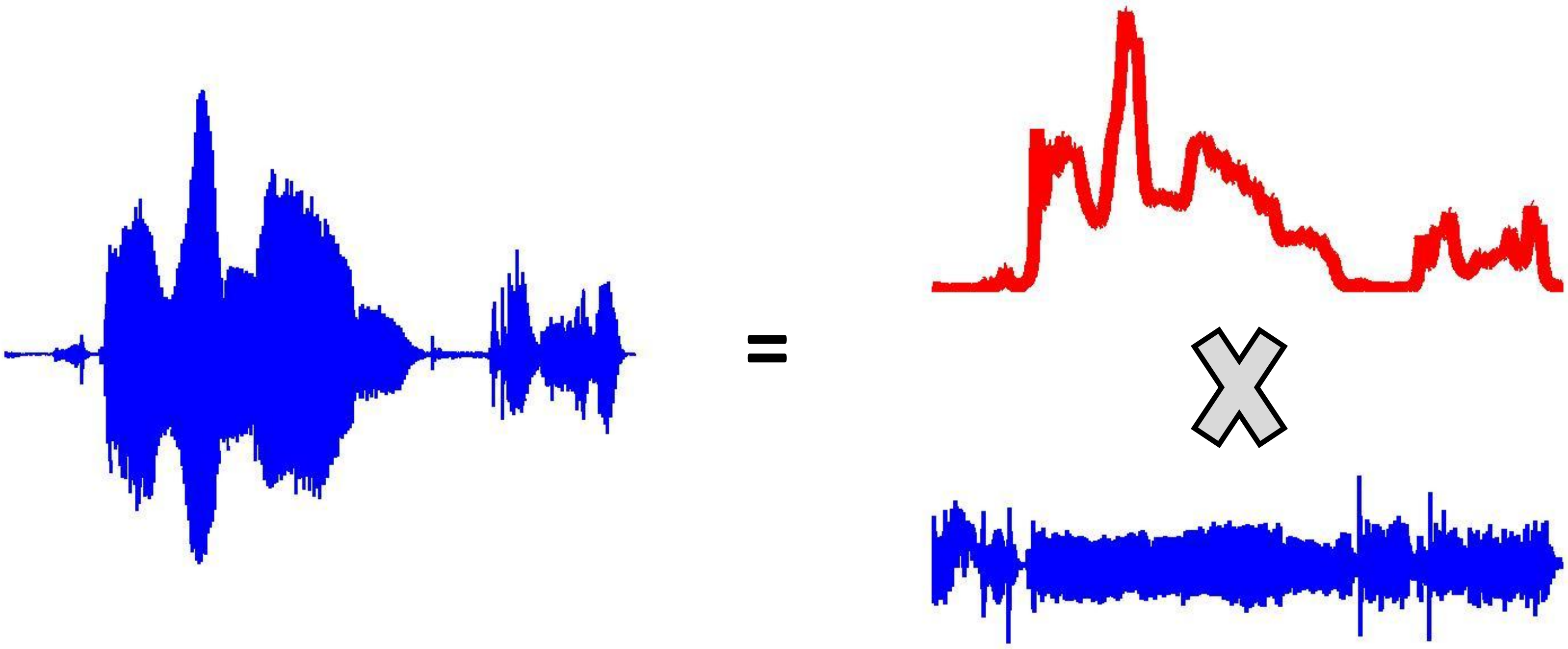
Introduction

- Sub-band speech and audio signals - product of smooth modulation with a fine carrier.



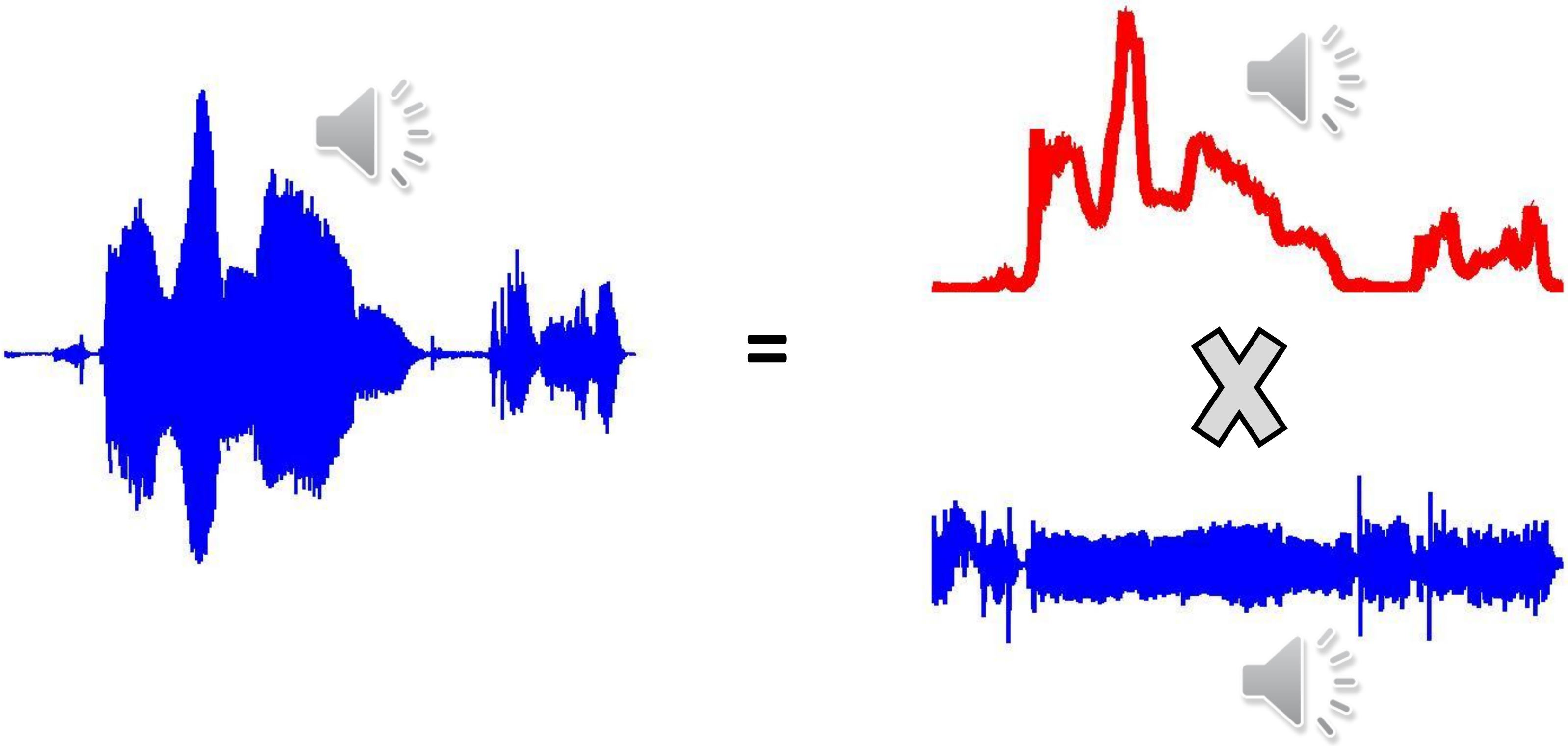
Introduction

- Sub-band speech and audio signals - product of smooth modulation with a fine carrier.



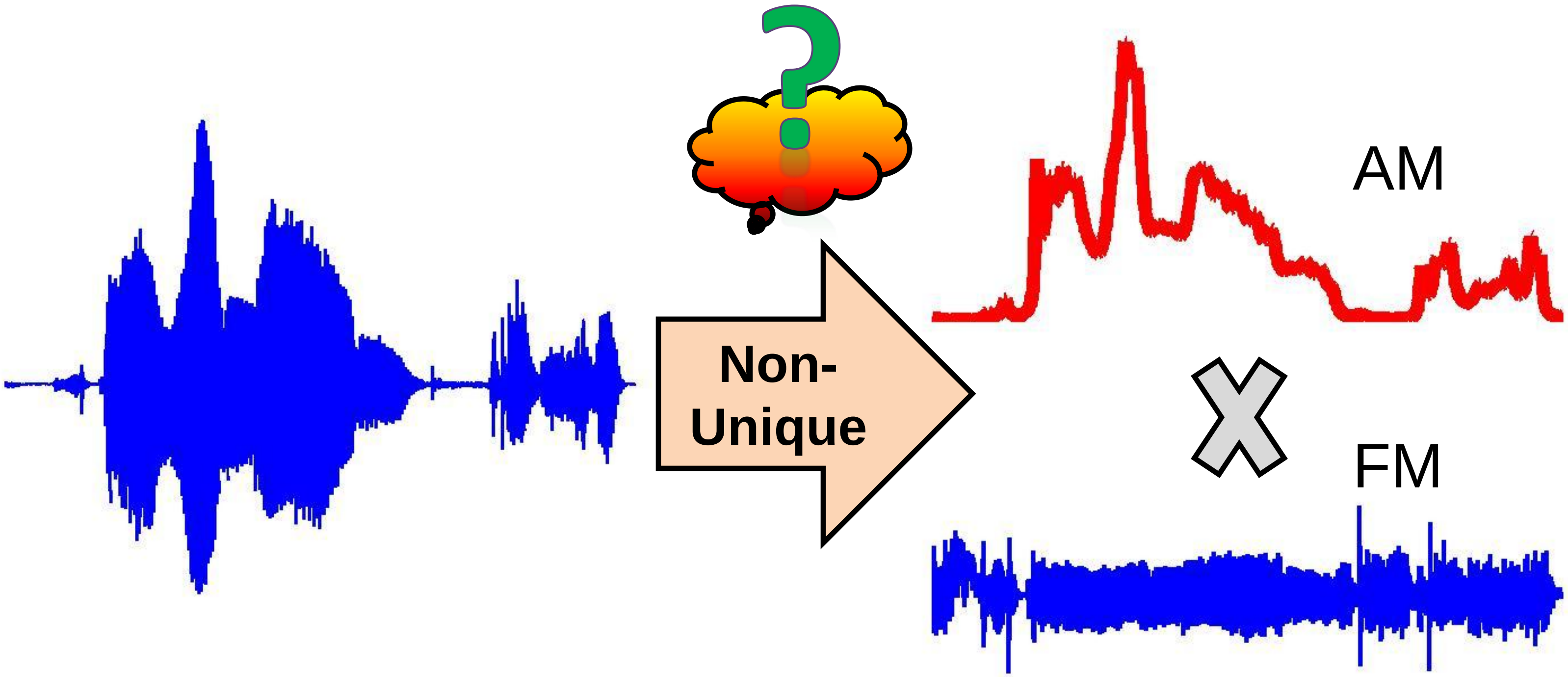
Introduction

- Sub-band speech and audio signals - product of smooth modulation with a fine carrier.



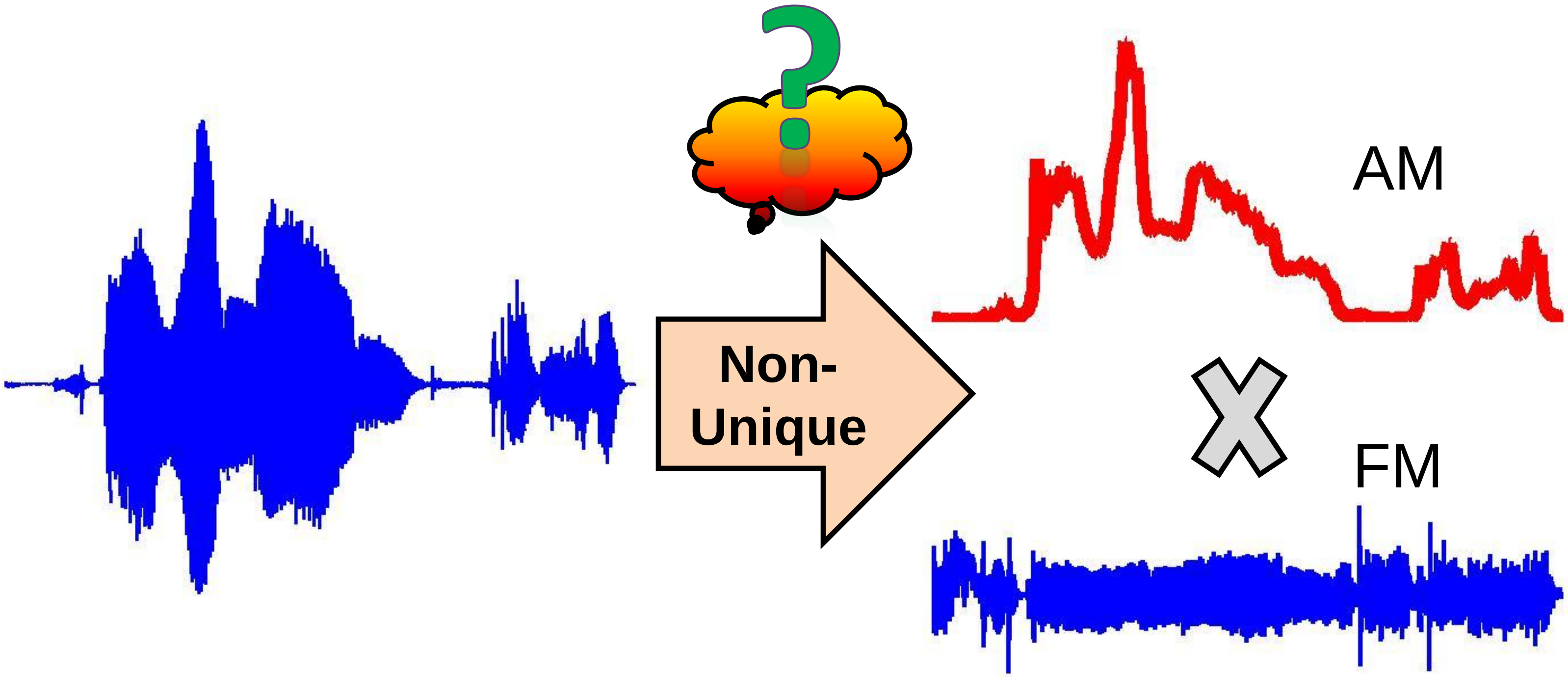
Introduction

- Sub-band speech and audio signals - product of smooth modulation with a fine carrier.



Introduction

- Sub-band speech and audio signals - product of smooth modulation with a fine carrier.



$$x(t) \quad = \quad m(t) * \cos\{\omega_o t + \varphi(t)\}$$

Desired Properties of AM

- Linearity

$$\alpha x(t) \quad \Longrightarrow \quad \alpha m(t)$$

- Continuity

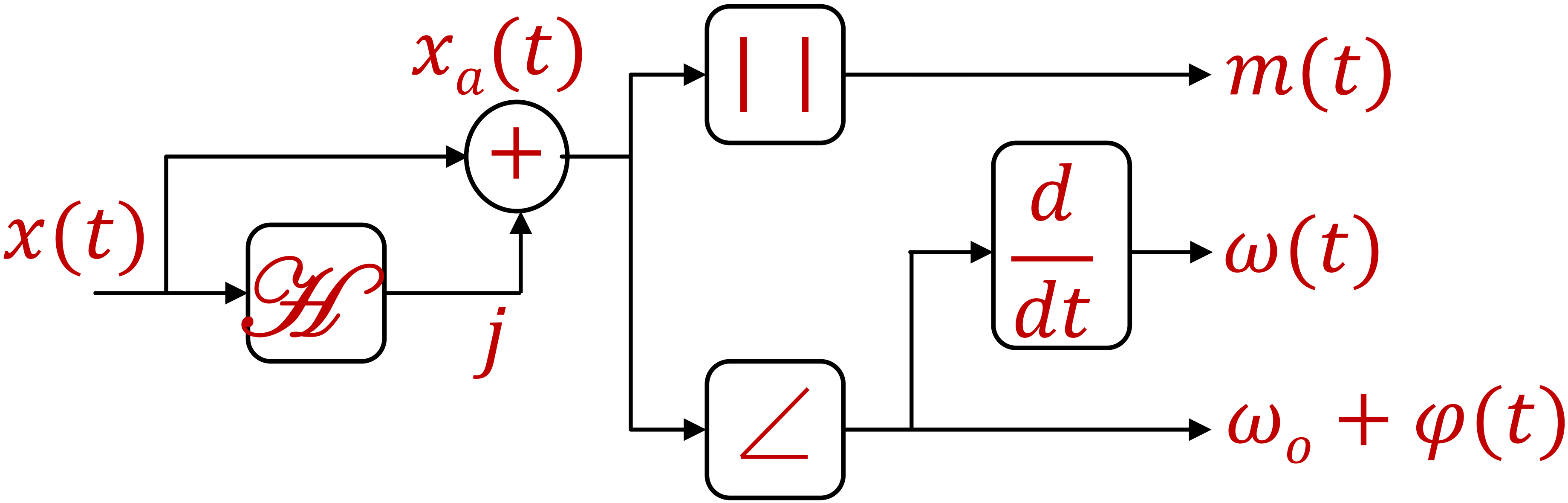
$$x(t) + \delta x(t) \quad \Longrightarrow \quad m(t) + \delta m(t)$$

- Harmonicity

$$\cos(\omega_o t) \quad \Longrightarrow \quad 1$$

Desired Properties of AM

- Uniquely satisfied by the analytic signal



$\mathscr{H}$ - Hilbert transform, $x_a(t)$ - analytic signal,

$|x_a(t)|^2$ – Hilbert envelope

Desired Properties of AM

- However, the Hilbert transform filter is infinitely long and can cause artifacts for finite length signals.

$$\mathscr{H}(x(t)) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x(t-\tau)}{t-\tau}\,d\tau$$

- Need for modeling the Hilbert envelope without explicit computation of the Hilbert transform.

Overview

- Introduction
- AR Model of Hilbert Envelopes
- FDLP and its Properties
- Applications
- Summary

Overview

- Introduction
- AR Model of Hilbert Envelopes
- FDLP and its Properties
- Applications
- Summary

AR Model of Hilbert Envelopes

Signal $x[n]$ with zero mean in time and frequency domain for $\mathrm{n} = 0...\mathrm{N}\text{-}1$

Discrete-time analytic spectrum

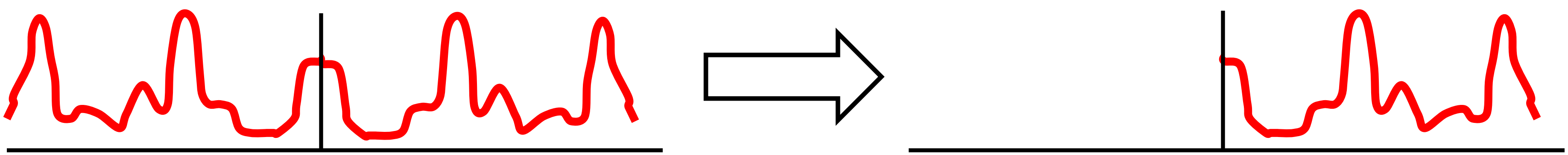
$$X_a[k] = \begin{cases} 2X[k] \text{ for } \mathrm{k} < \mathrm{N}/2 \\ 0 \qquad\quad \text{for } \mathrm{k} \geq \mathrm{N}/2 \end{cases}$$

AR Model of Hilbert Envelopes

Signal $x[n]$ with zero mean in time and frequency domain for $\mathrm{n} = 0 \ldots \mathrm{N}\text{-}1$

Discrete-time analytic spectrum

$$X_a[k] = \begin{cases} 2X[k] & \text{for } \mathrm{k} < \mathrm{N}/2 \\ 0 & \text{for } \mathrm{k} \geq \mathrm{N}/2 \end{cases}$$



$X[k]$ $X_a[k]$

AR Model of Hilbert Envelopes

Let $q[n]$ - even-symmetrized version of $x[n]$.

$q[n] = x[n] \text{ for n} < \text{N}, \quad q[n] = x[M-n], M = 2N-1$

Spectrum

$$Q[k] = 2Re\{X[k]\}$$

AR Model of Hilbert Envelopes

Let $q[n]$ - even-symmetrized version of $x[n]$.

$q[n] = x[n]$ for n < N, $q[n] = x[M-n], M = 2N-1$

Discrete-time analytic spectrum

$$Q[k] = 2Re\{X[k]\}$$

$$Q_a[k] = \begin{cases} 2Q[k], & \mathrm{k<N} \\ 0 & \mathrm{k \geq N} \end{cases}$$

AR Model of Hilbert Envelopes

Let $q[n]$ - even-symmetrized version of $x[n]$.

$q[n] = x[n]$ for n < N, $q[n] = x[M - n], M = 2N - 1$

Discrete-time analytic spec.

$$Q[k] = 2Re\{X[k]\}$$

$$Q_a[k] = \begin{cases} 2Q[k], & \text{k<N} \\ 0 & \text{k} \geq \text{N} \end{cases}$$

N-point DCT

$$y[k] = 4Re\{X[k]\}, \text{k<N}$$

AR Model of Hilbert Envelopes

Let $q[n]$ - even-symmetrized version of $x[n]$.

$q[n] = x[n]$ for n < N, $q[n] = x[M - n], M = 2N - 1$

Discrete-time analytic spec.

$$Q[k] = 2Re\{X[k]\}$$

$$Q_a[k] = \begin{cases} 2Q[k], & \text{k<N} \\ 0 & \text{k} \geq \text{N} \end{cases}$$

DCT zero-padded with N-zeros

$$\widehat{y[k]} = \begin{cases} 4Re\{X[k]\} & \text{k<N} \\ 0 & \text{k} \geq \text{N} \end{cases}$$

AR Model of Hilbert Envelopes

Let $q[n]$ - even-symmetrized version of $x[n]$.

$q[n] = x[n]$ for n < N, $q[n] = x[M-n], M = 2N-1$

Discrete-time analytic spec.

$$Q[k] = 2Re\{X[k]\}$$

$$Q_a[k] = \begin{cases} 2Q[k], & \text{k<N} \\ 0 & \text{k} \geq \text{N} \end{cases}$$

DCT zero-padded with N-zeros

$$\widehat{y[k]} = \begin{cases} 4Re\{X[k]\} & \text{k<N} \\ 0 & \text{k} \geq \text{N} \end{cases}$$

$$Q_a[k] = \mathscr{F}\{q_a[n]\} = \widehat{y[k]}$$

AR Model of Hilbert Envelopes

We have shown -

$$Q_a[k] = \mathscr{F}\{q_a[n]\} = \widehat{y[k]}$$

Even-sym. analytic spectrum.

Zero-padded DCT sequence

AR Model of Hilbert Envelopes

We have shown -

$$Q_a[k] = \mathscr{F}\{q_a[n]\} = \widehat{y[k]}$$

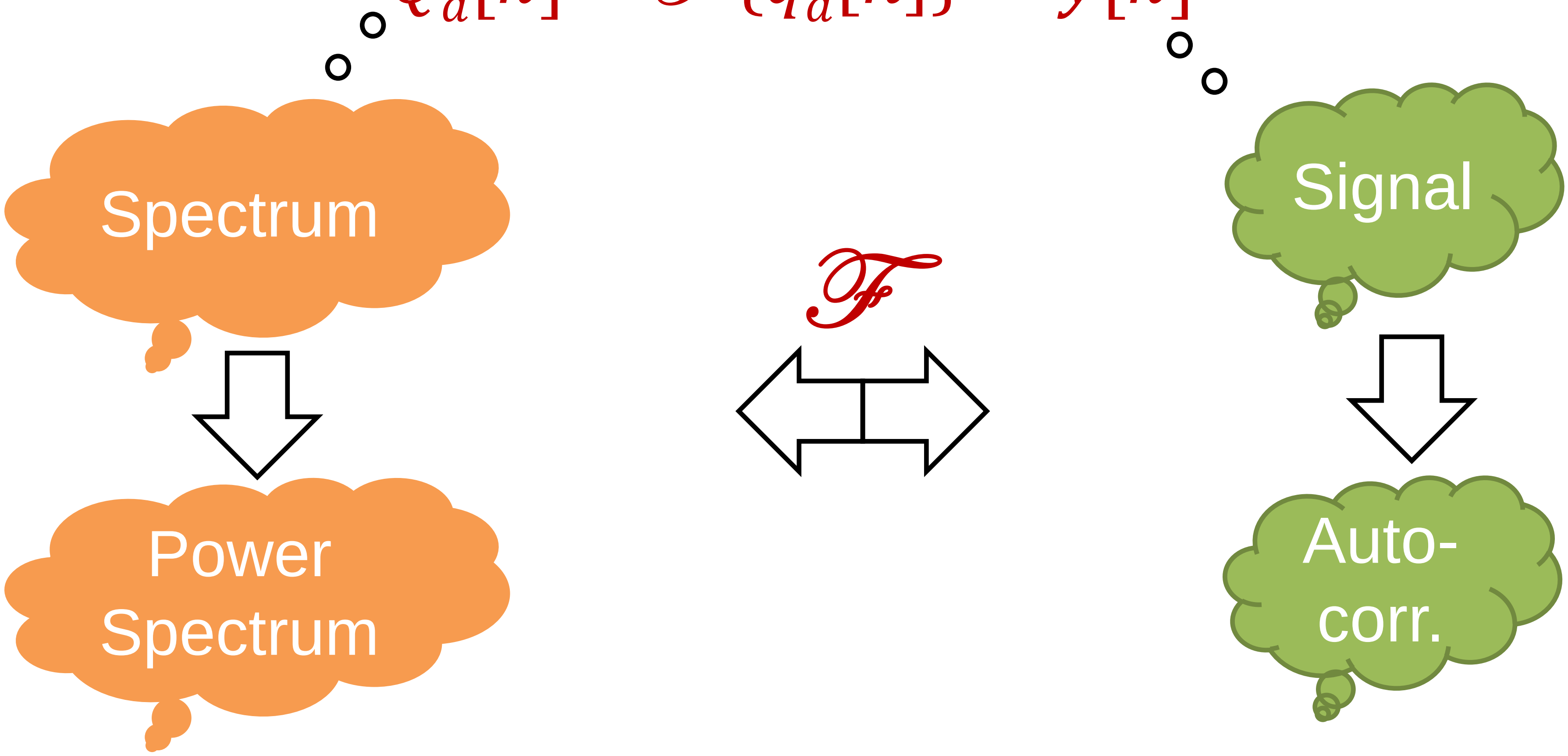
Spectrum

Signal

AR Model of Hilbert Envelopes

We have shown -

$$Q_a[k] = \mathscr{F}\{q_a[n]\} = \widehat{y[k]}$$



AR Model of Hilbert Envelopes

We have shown -

$$Q_a[k] = \mathscr{F}\{q_a[n]\} = \widehat{y[k]}$$

Even-sym. analytic spectrum.

Zero-padded DCT sequence

AR Model of Hilbert Envelopes

We have shown -

$$Q_a[k] = \mathscr{F}\{q_a[n]\} = \widehat{y[k]}$$

$$\boxed{\mathscr{F}\{|q_a[n]|^2\} = r_y[\tau]}$$

Spectrum of Hilbert env. for even-sym. signal

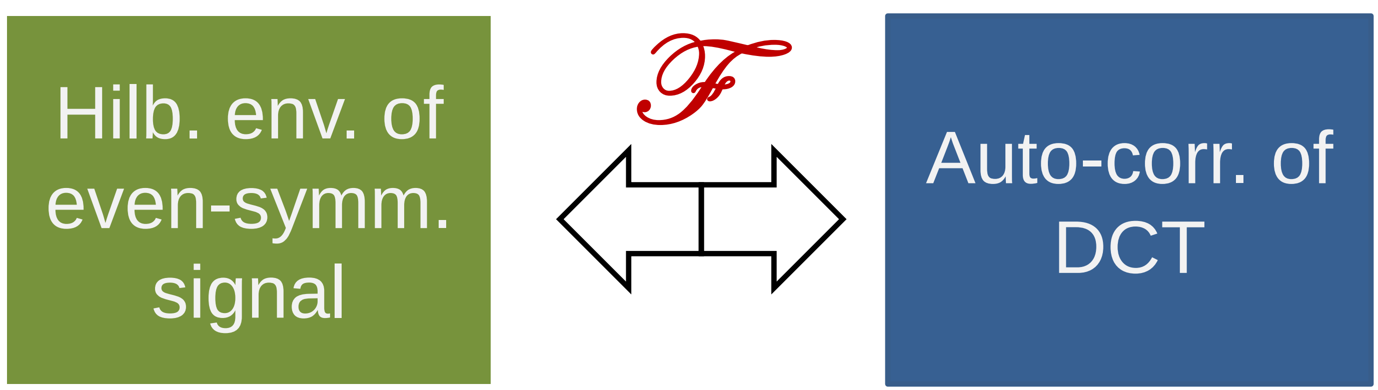
Auto-correlation of DCT sequence

AR Model of Hilbert Envelopes

We have shown -

$$Q_a[k] = \mathscr{F}\{q_a[n]\} = \widehat{y[k]}$$

$$\boxed{\mathscr{F}\{|q_a[n]|^2\} = r_y[\tau]}$$



AR Model of Hilbert Envelopes

We have shown -

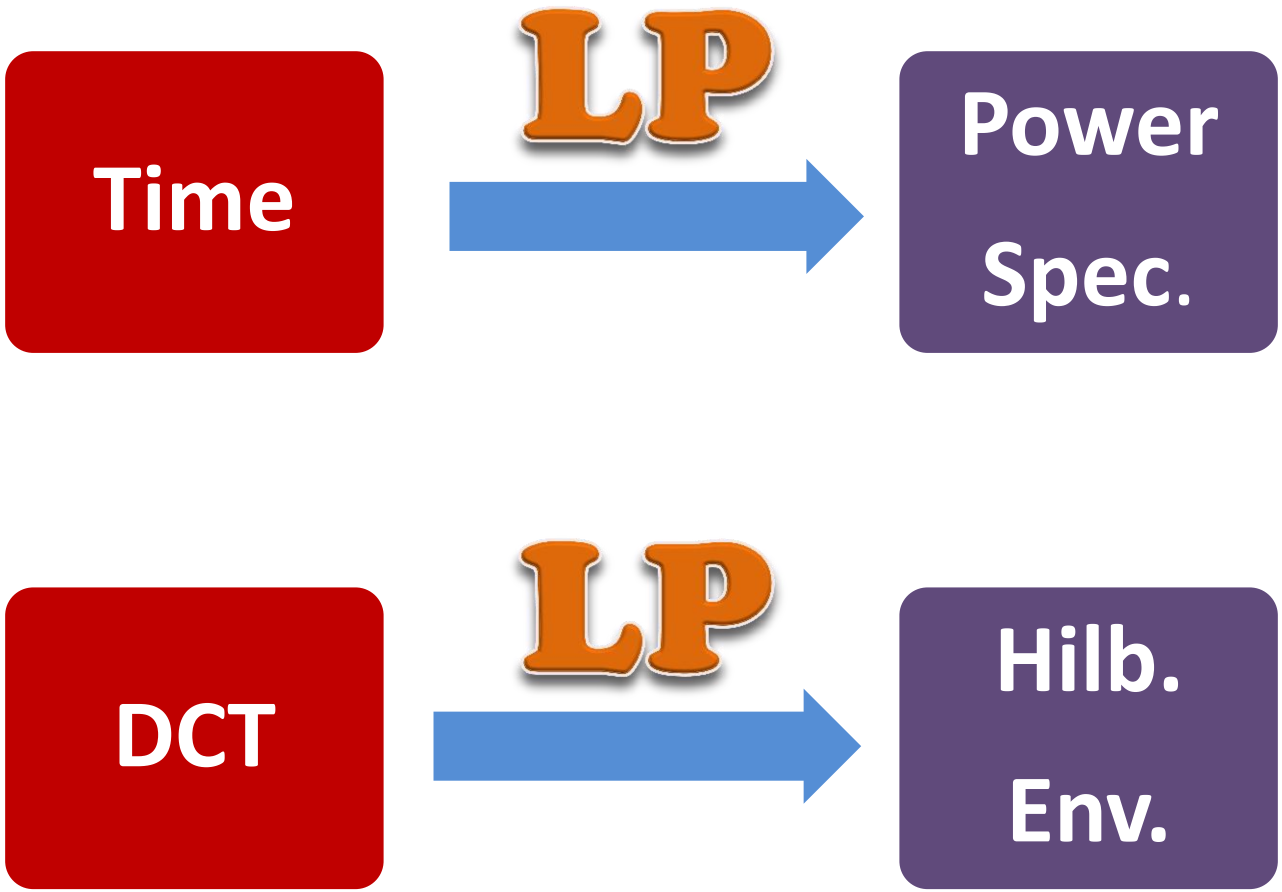
$$Q_a[k] = \mathscr{F}\{q_a[n]\} = \widehat{y[k]}$$

$$\boxed{\mathscr{F}\{|q_a[n]|^2\} = r_y[\tau]}$$

AR model of Hilb. env. $\overset{\textbf{LP}}{\Longleftrightarrow}$ Auto-corr. of DCT

LP in Time and Frequency

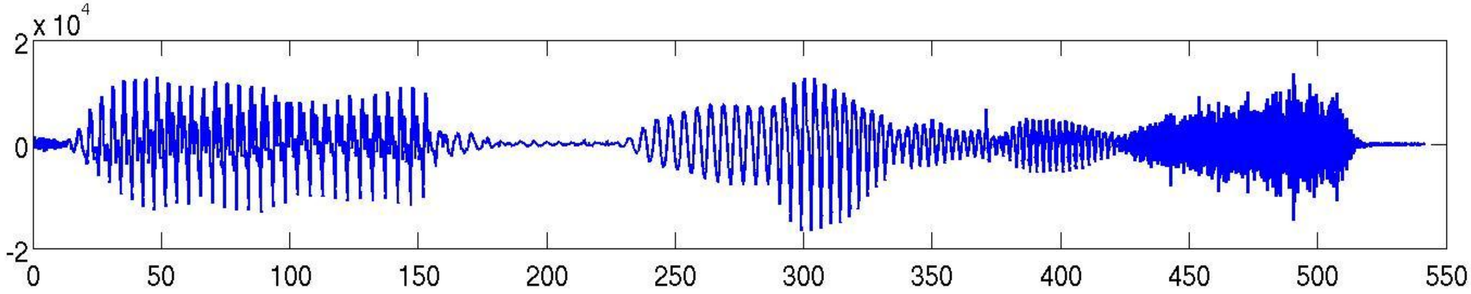
LP in Time and Frequency



FDLP

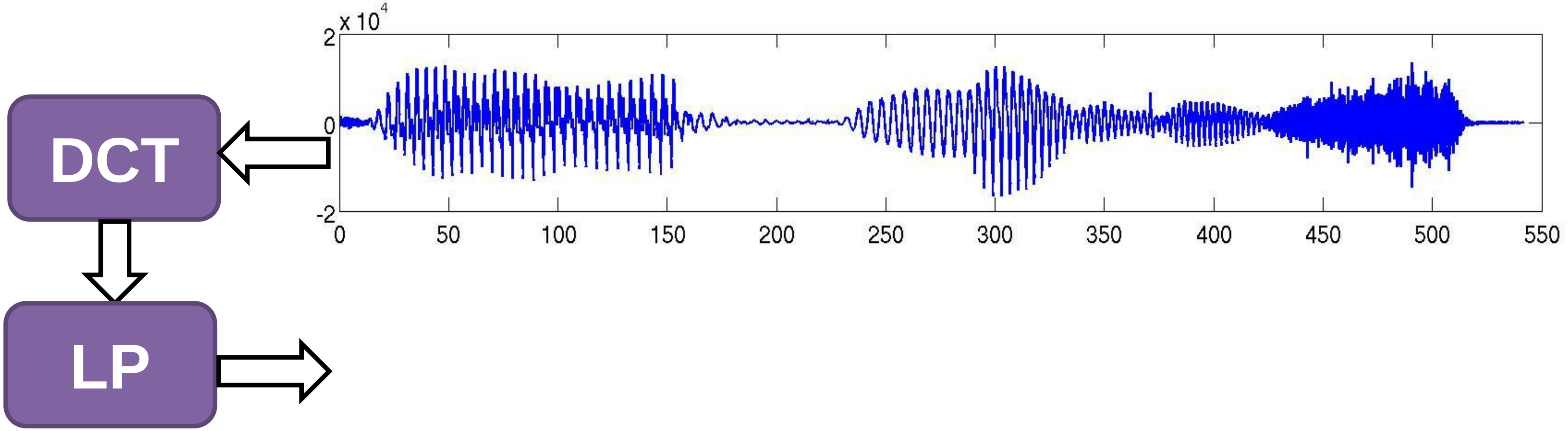
Linear prediction on the cosine transform of the signal

Speech



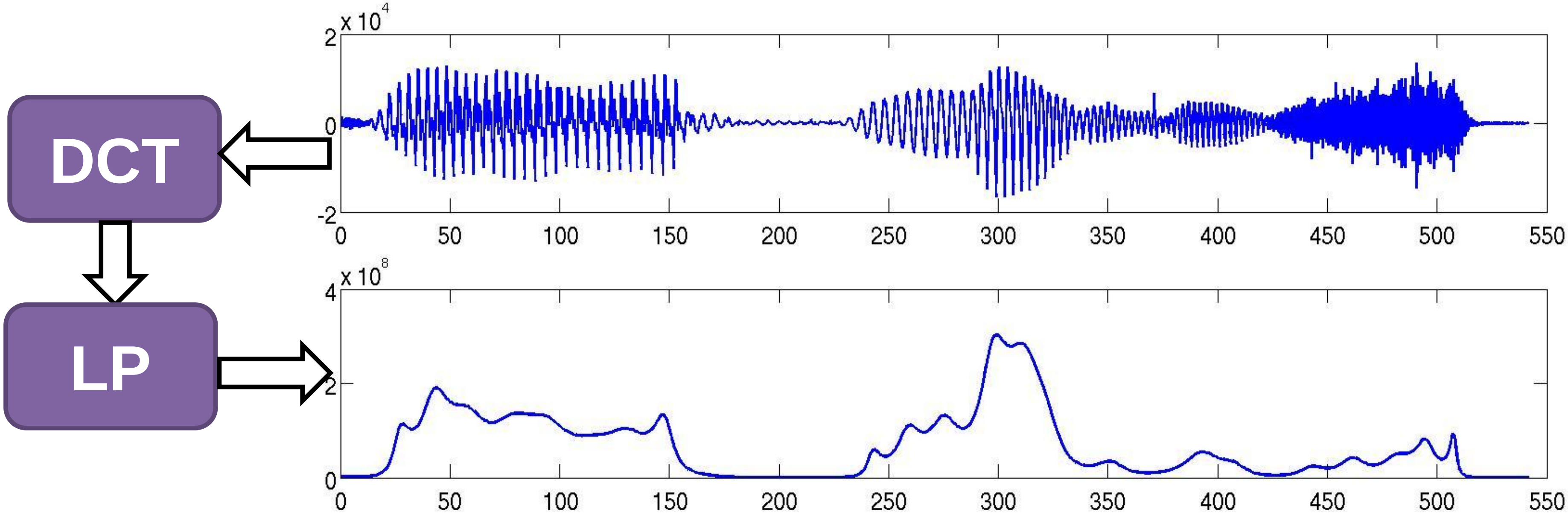
FDLP

Linear prediction on the cosine transform of the signal



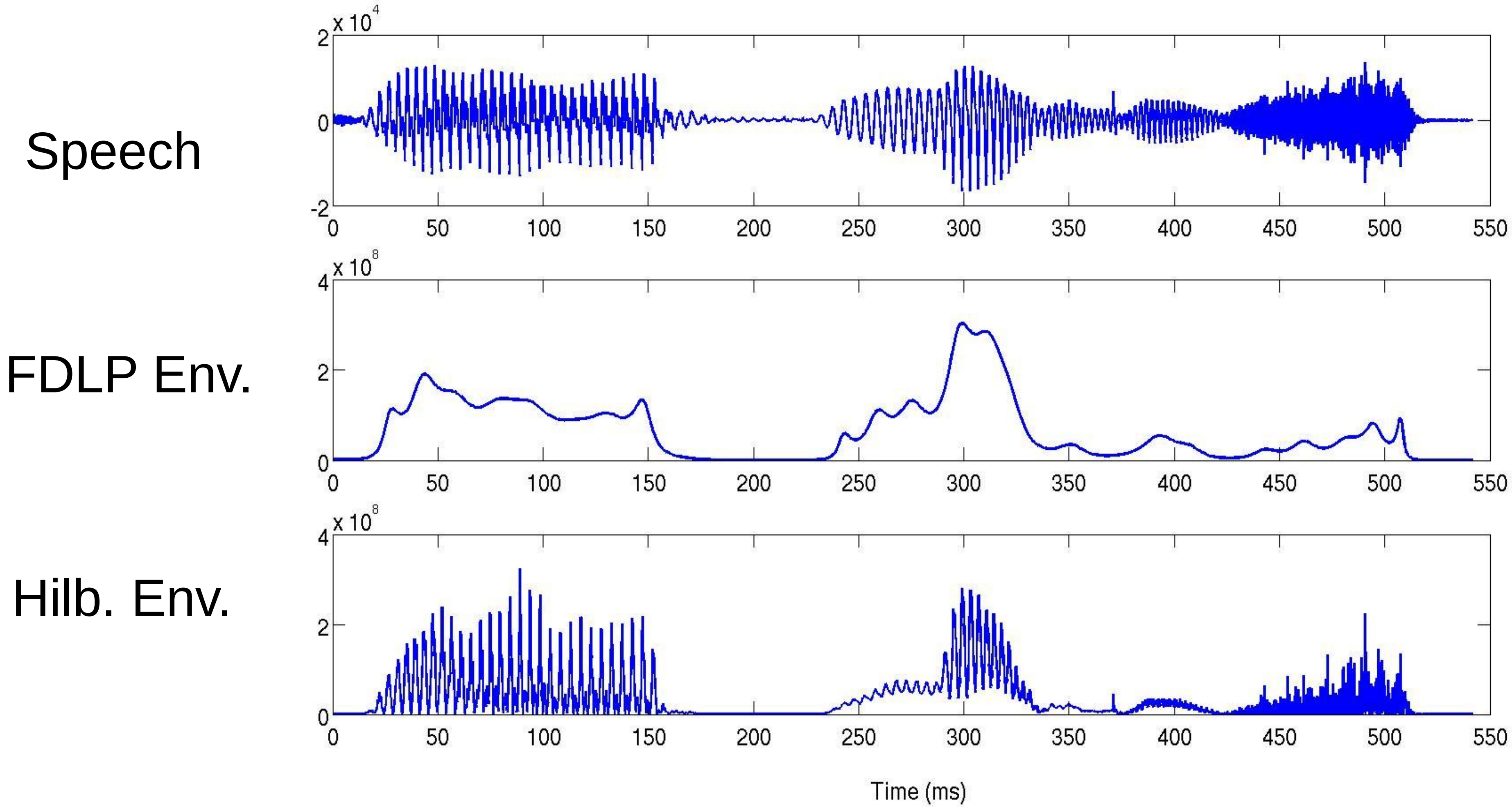
FDLP

Linear prediction on the cosine transform of the signal



FDLP

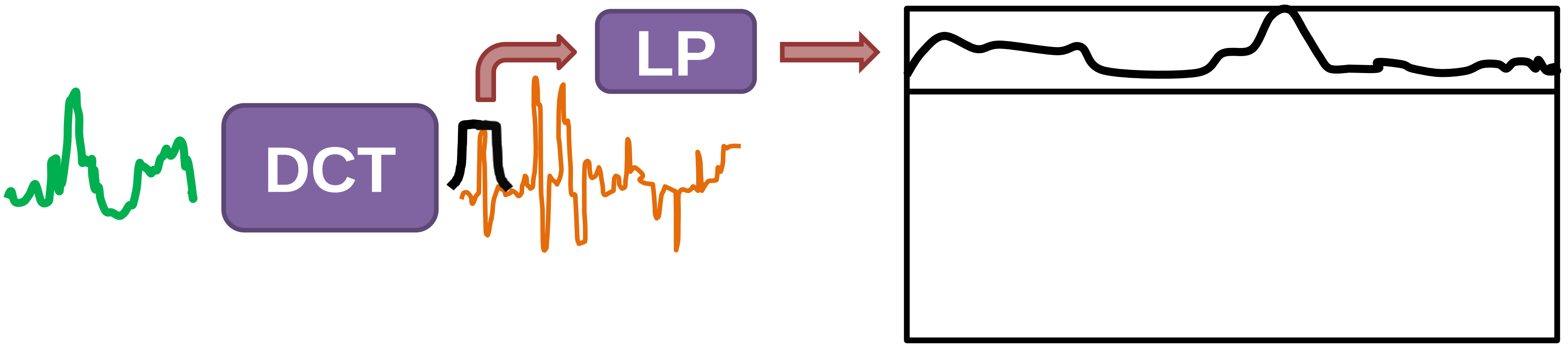
Linear prediction on the cosine transform of the signal



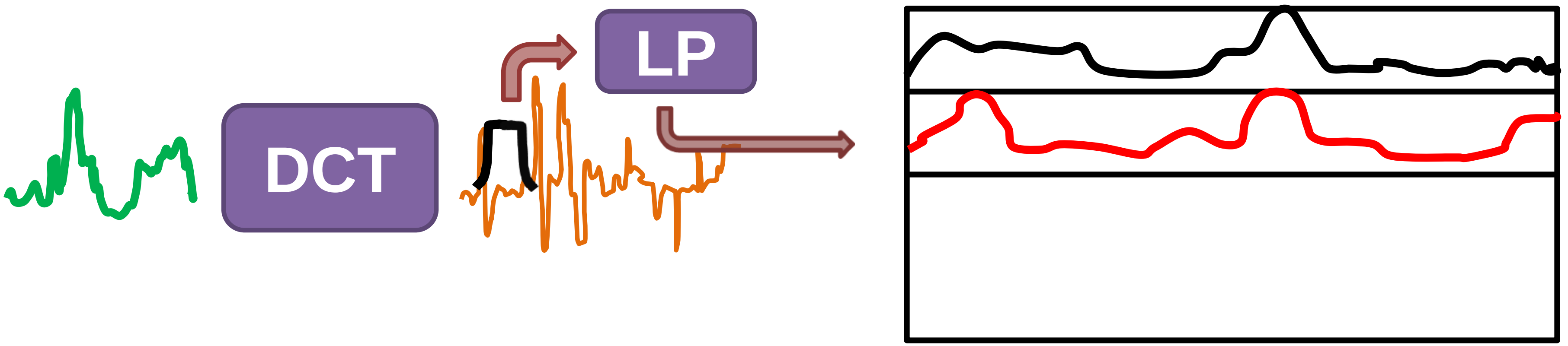
FDLP for Speech Representation

FDLP for Speech Representation

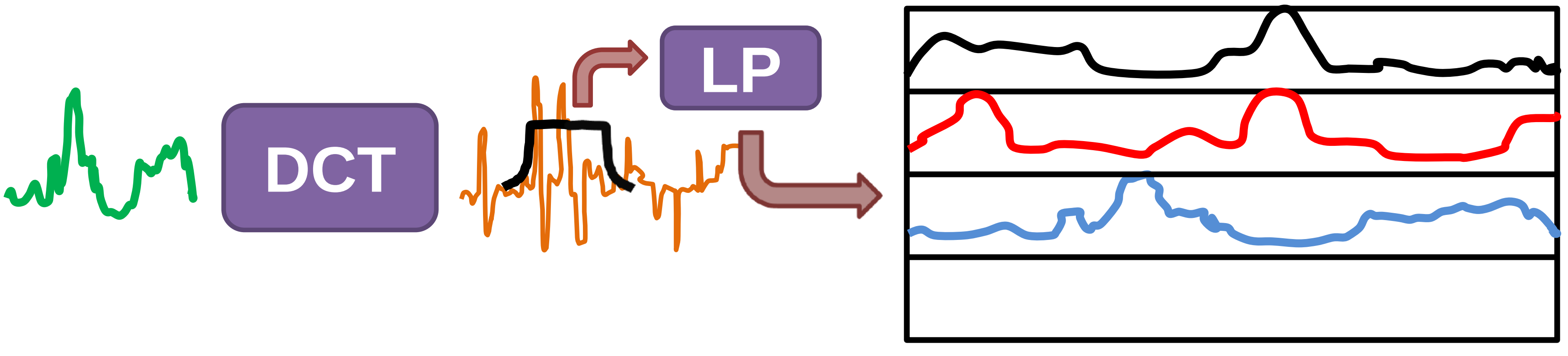
FDLP for Speech Representation



FDLP for Speech Representation



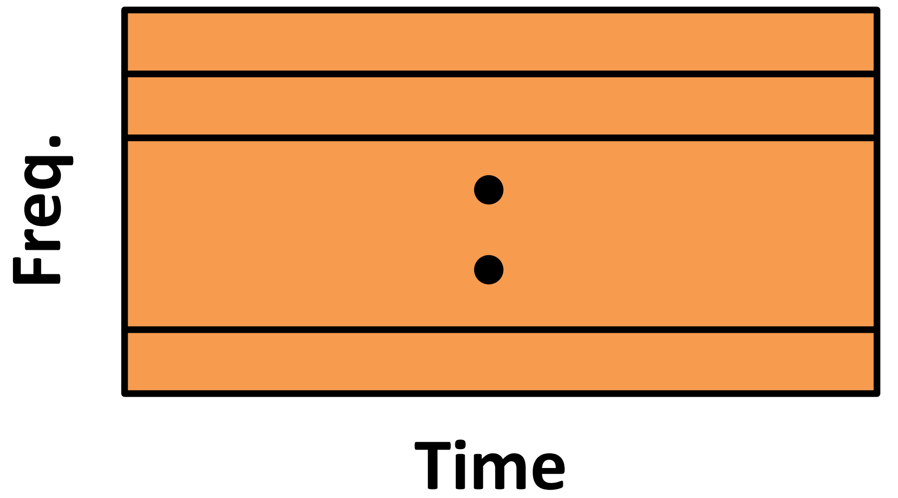
FDLP for Speech Representation



FDLP for Speech Representation

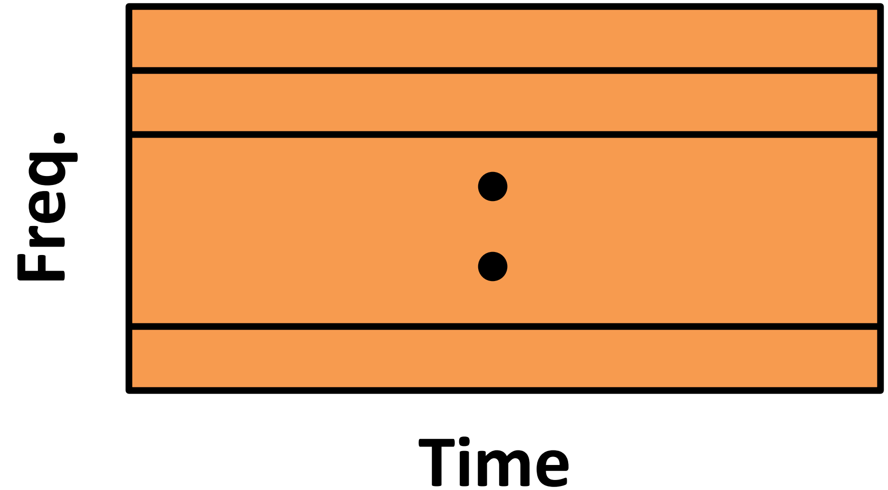
FDLP

Spectrogram

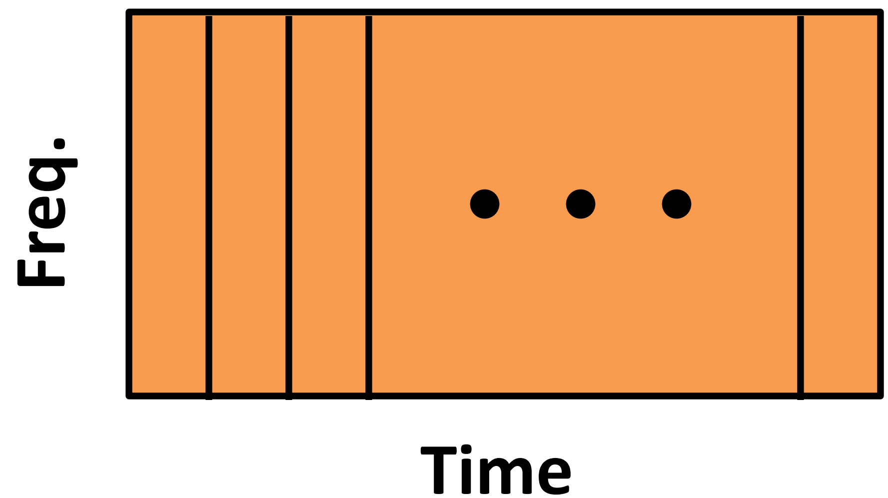


FDLP for Speech Representation

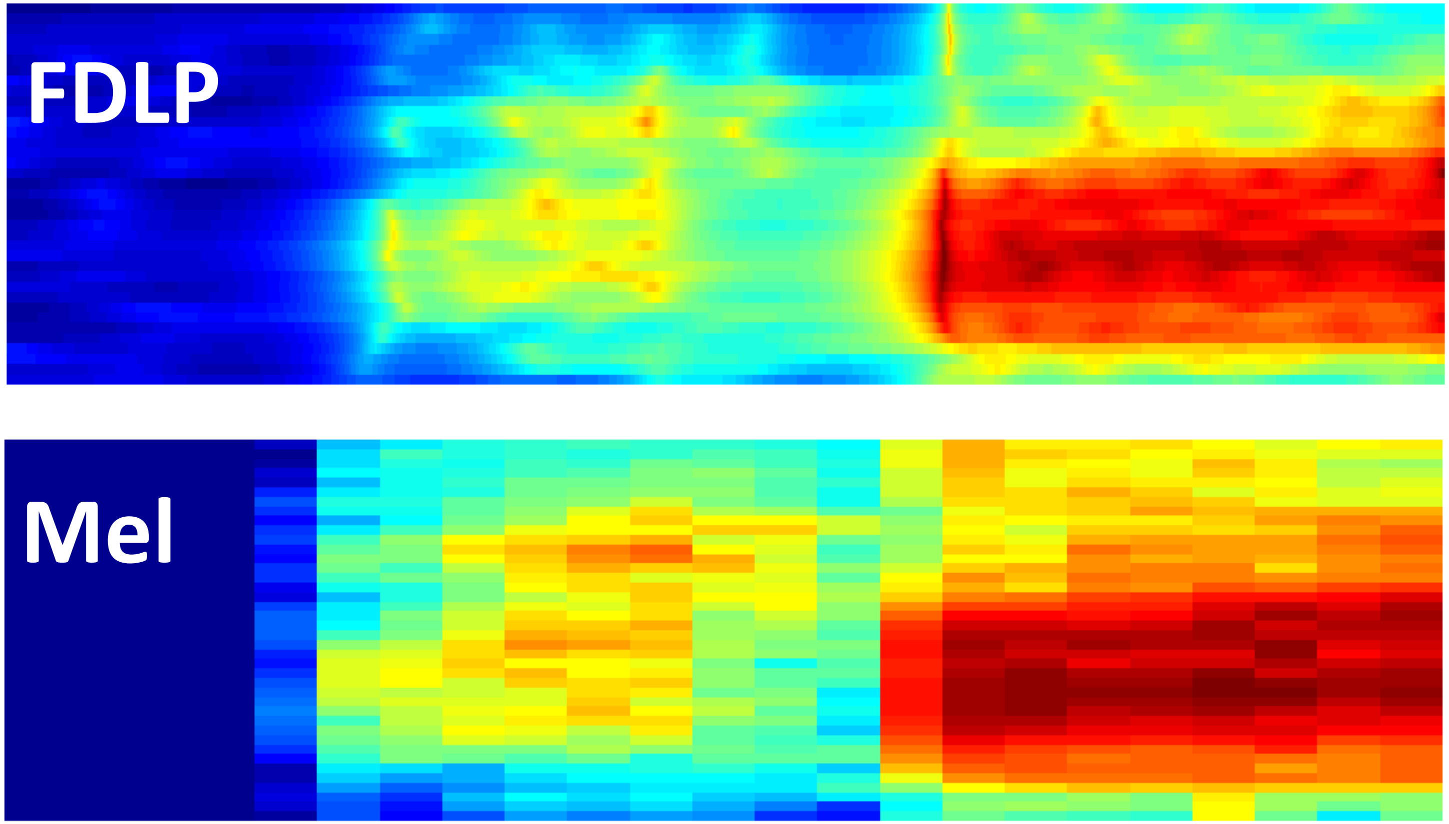
FDLP Spectrogram



Conventional Approaches



FDLP versus Mel Spectrogram



Sriram Ganapathy, Samuel Thomas and H. Hermansky, “Comparison of Modulation Frequency Features for Speech Recognition", *ICASSP, 2010.*

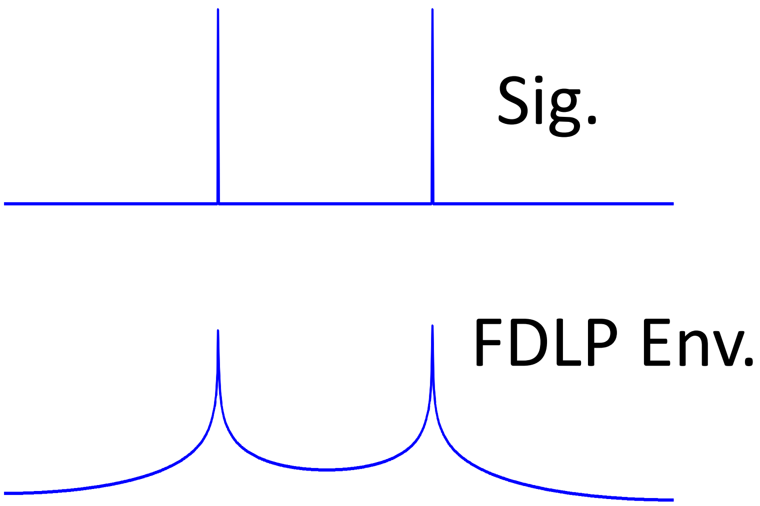
Overview

- Introduction
- AR Model of Hilbert Envelopes
- FDLP and its Properties
- Applications
- Summary

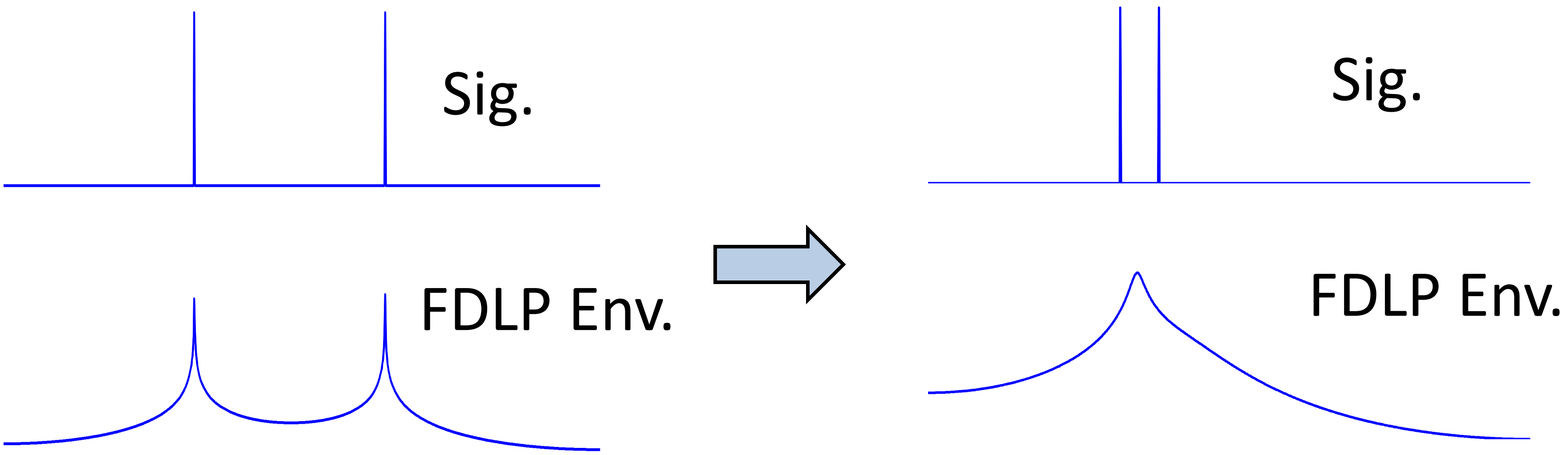
Overview

- Introduction
- AR Model of Hilbert Envelopes
- FDLP and its Properties
- Applications
- Summary

Resolution of FDLP Analysis

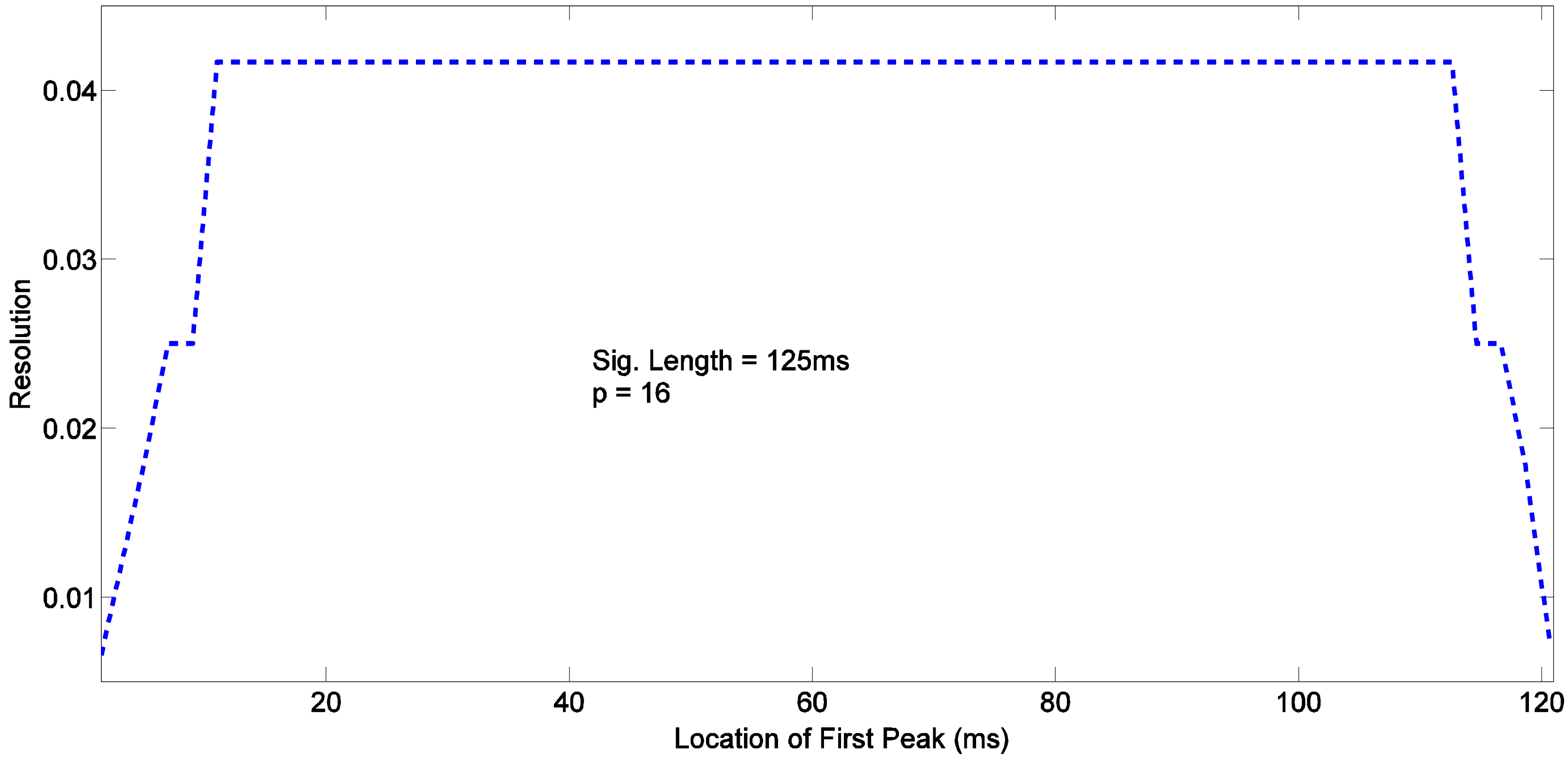


Resolution of FDLP Analysis

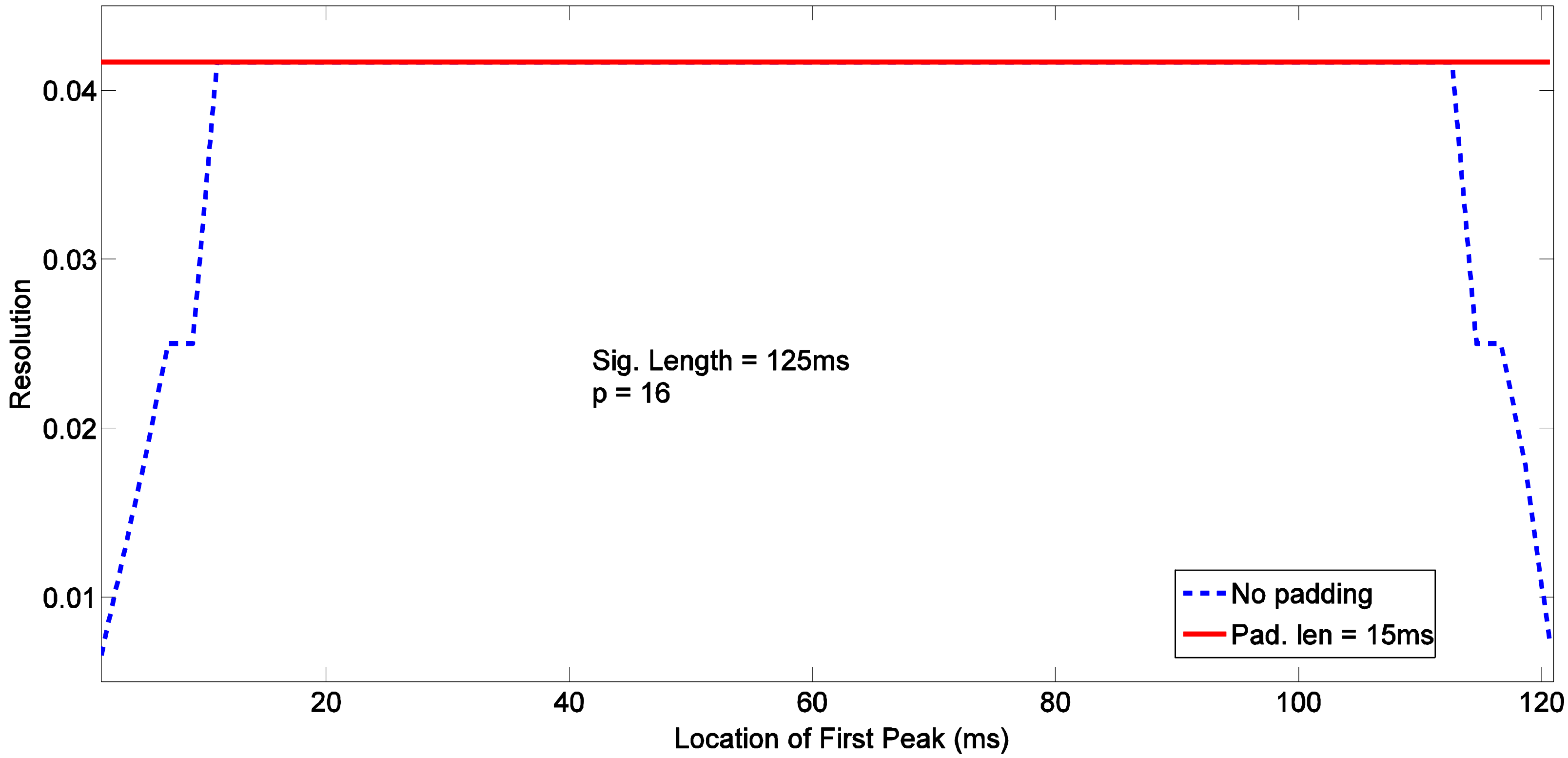


Res. = (Critical Width)$^{-1}$

Resolution of FDLP Analysis



Resolution of FDLP Analysis



Properties of FDLP Analysis

- Summarizing the gross temporal variation with a few parameters
  - Model order of FDLP controls the degree of smoothness.
  - AR model captures perceptually important high energy regions of the signal.

- Suppressing reverberation artifacts
  - Reverberation is a long-term convolutive distortion.
    - Analysis in long-term windows and narrow sub-bands.

Properties of FDLP Analysis

- Summarizing the gross temporal variation with a few parameters
  - Model order of FDLP controls the degree of smoothness.
  - AR model captures perceptually important high energy regions of the signal.

- Suppressing reverberation artifacts
  - Reverberation is a long-term convolutive distortion.
    - Analysis in long-term windows and narrow sub-bands.

Reverberation

When speech is corrupted with convolutive distortion like room reverberation

Clean Speech * Room Response = Revb. Speech

Reverberation

When speech is corrupted with convolutive distortion like room reverberation

Clean Speech * Room Response = Revb. Speech

In the long-term DFT domain, this translates

Clean DFT x Response DFT = Revb. DFT

Reverberation

When speech is corrupted with convolutive distortion like room reverberation

$$r[n] = x[n] * h[n]$$

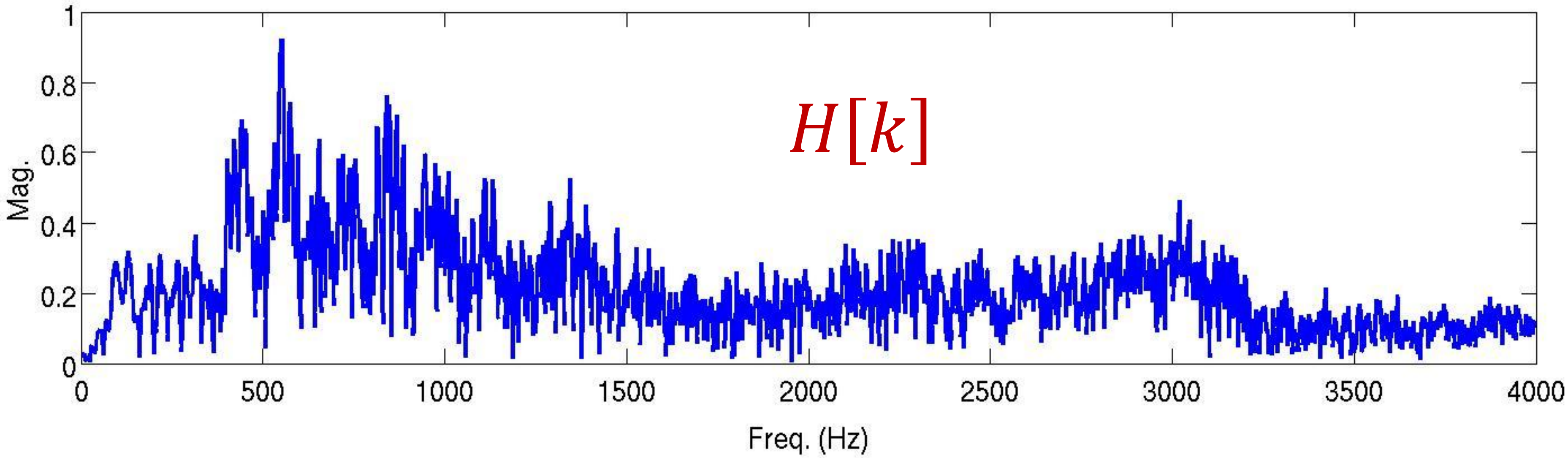
In the DFT domain, this translates to a multiplication

$$R[k] = X[k] \times H[k]$$

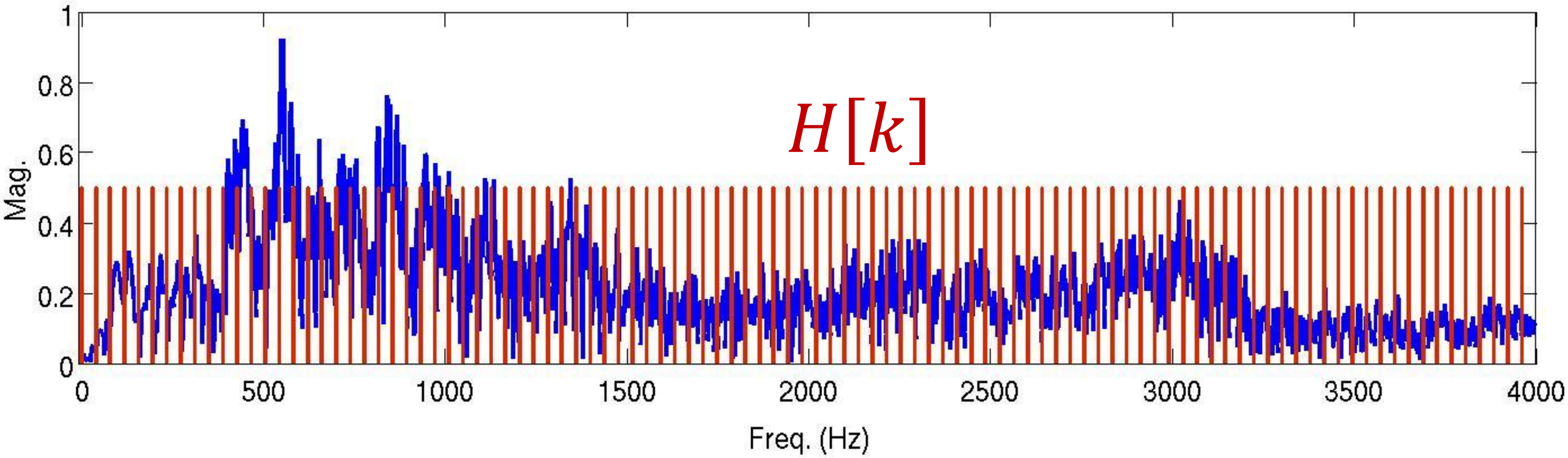
In the m^{th} sub-band,

$$R_m[k] = X_m[k] \times H_m[k]$$

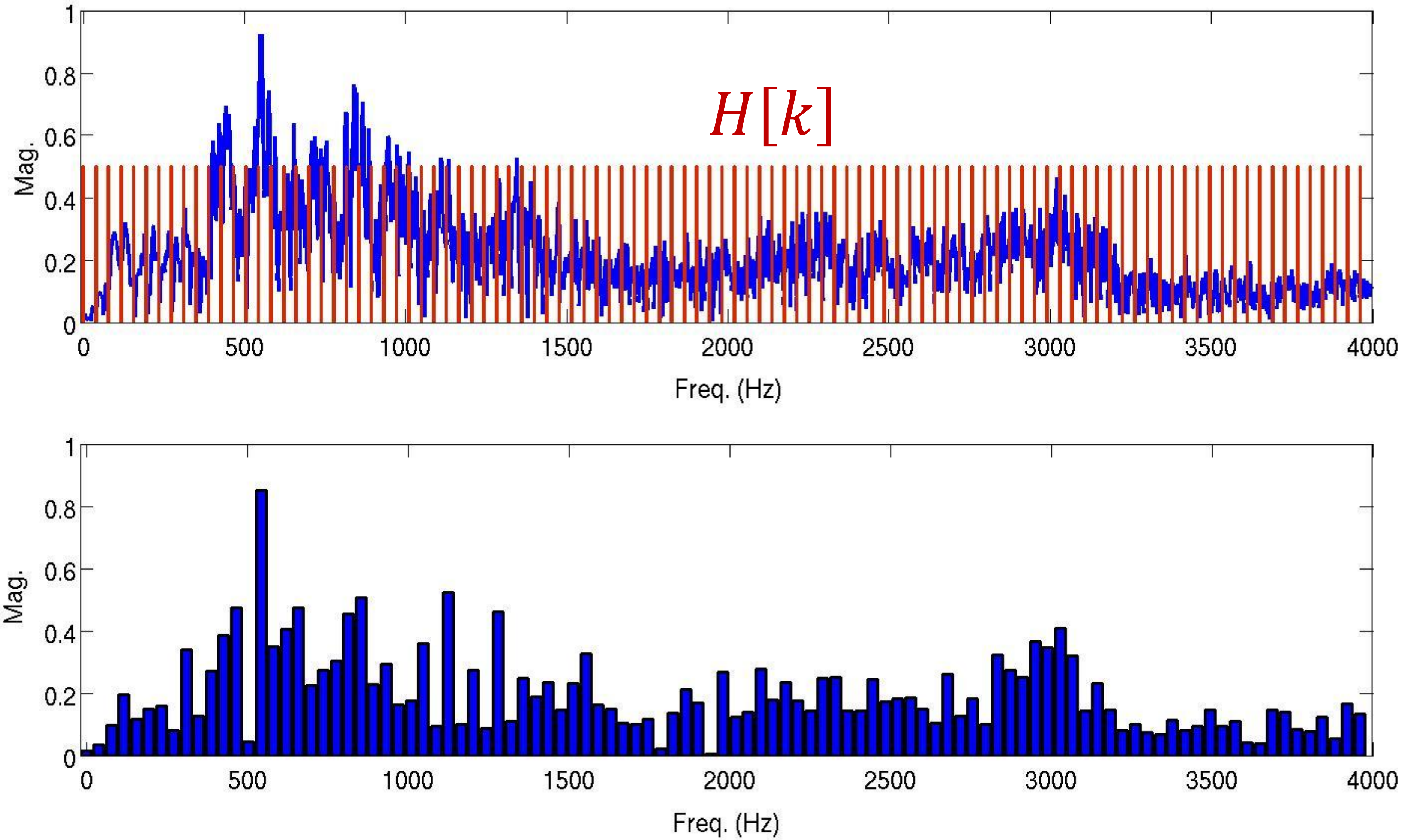
Reverberation



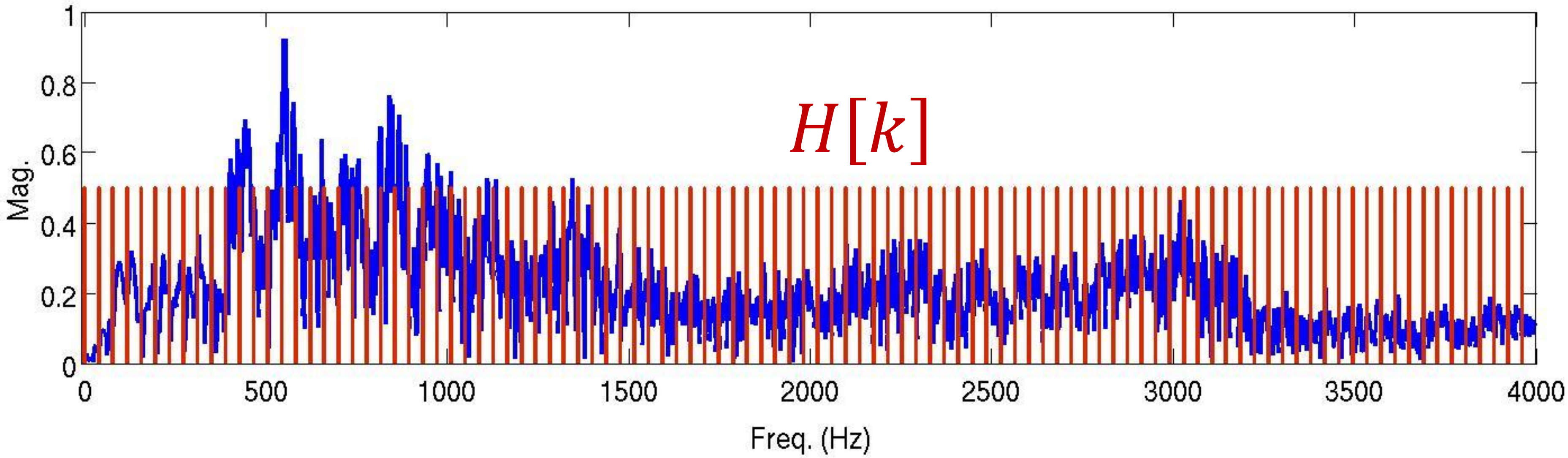
Reverberation

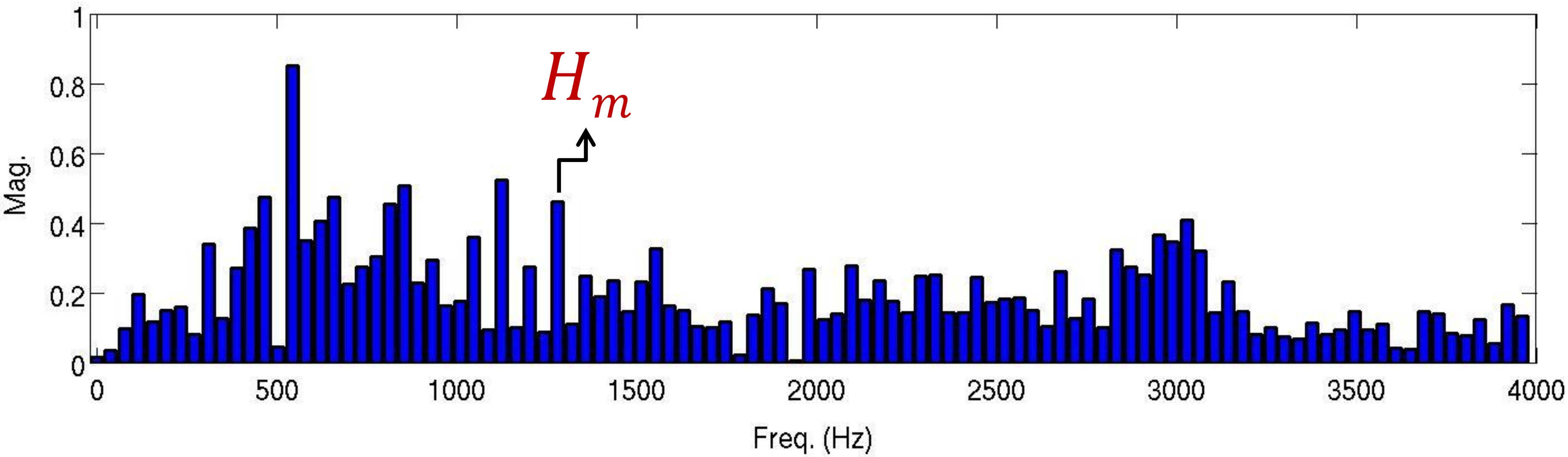


Reverberation



Reverberation





Reverberation

When speech is corrupted with convolutive distortion like room reverberation

$$r[n] = x[n] * h[n]$$

In the DFT domain, this translates to a multiplication

$$R[k] = X[k] \times H[k]$$

In the m^{th} sub-band,

$$R_m[k] = X_m[k] \times H_m[k]$$

In narrow bands, $H_m[k]$ is approx. constant,

$$R_m[k] \cong X_m[k] \times H_m$$

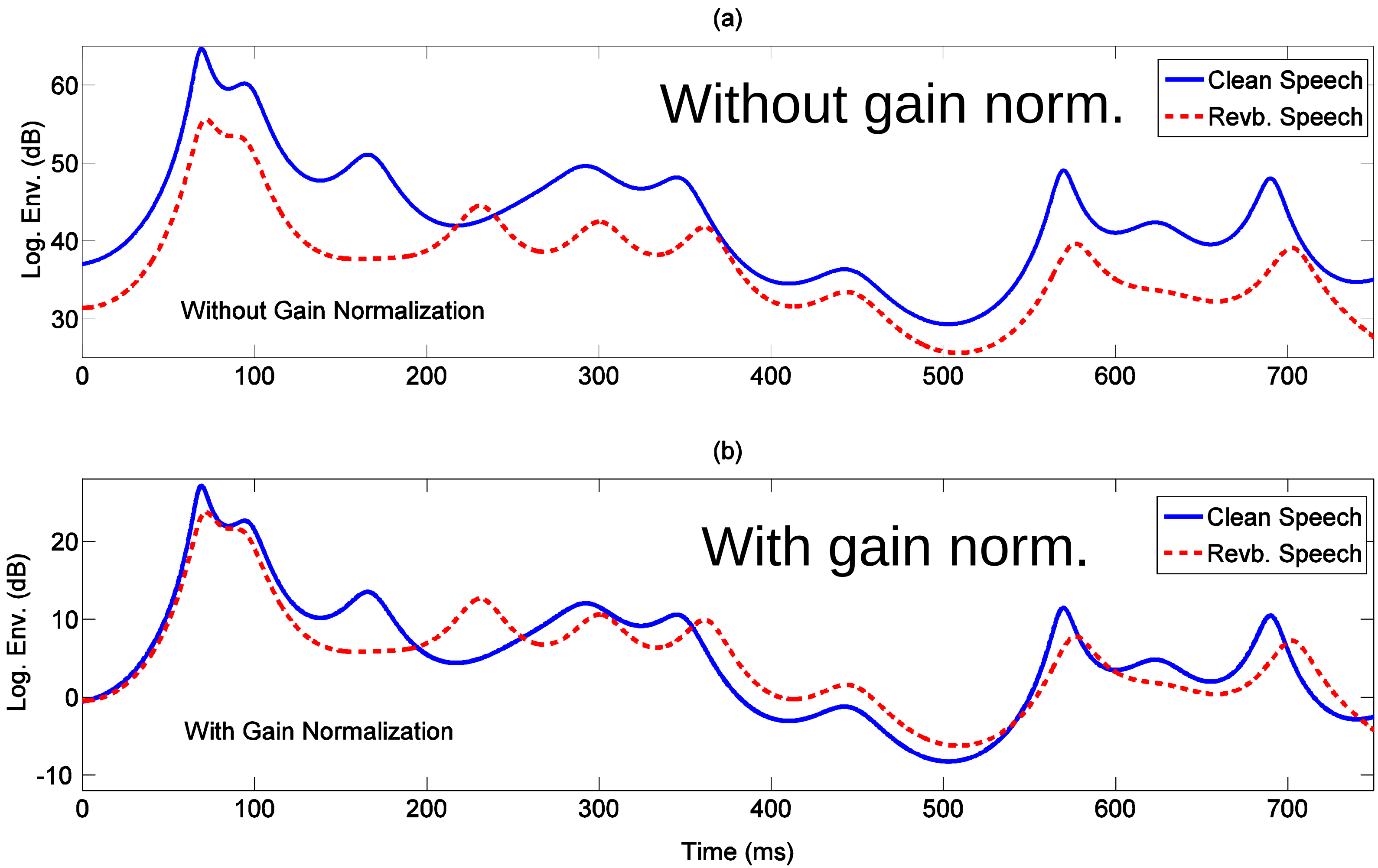
Gain Normalization in FDLP

- FDLP envelope of m^{th} band using all-pole parameters $\{a_1, \ldots a_p\}$ is given by

$$\widehat{E_m}[n] = \frac{G}{|1 - \sum_{k=1}^{p} a_k e^{\frac{-j2\pi kn}{N}}|^2}$$

- When the sub-band signal is multiplied by H_m, the gain G is modified.
- Normalization to convolutive distortions is achieved by reconstructing the FDLP envelope with $G = 1$.

Gain Normalization in FDLP



S. Thomas, S. Ganapathy and H. Hermansky, “Recognition of Reverberant Speech Using FDLP", *IEEE Signal Proc. Letters, 2008.*

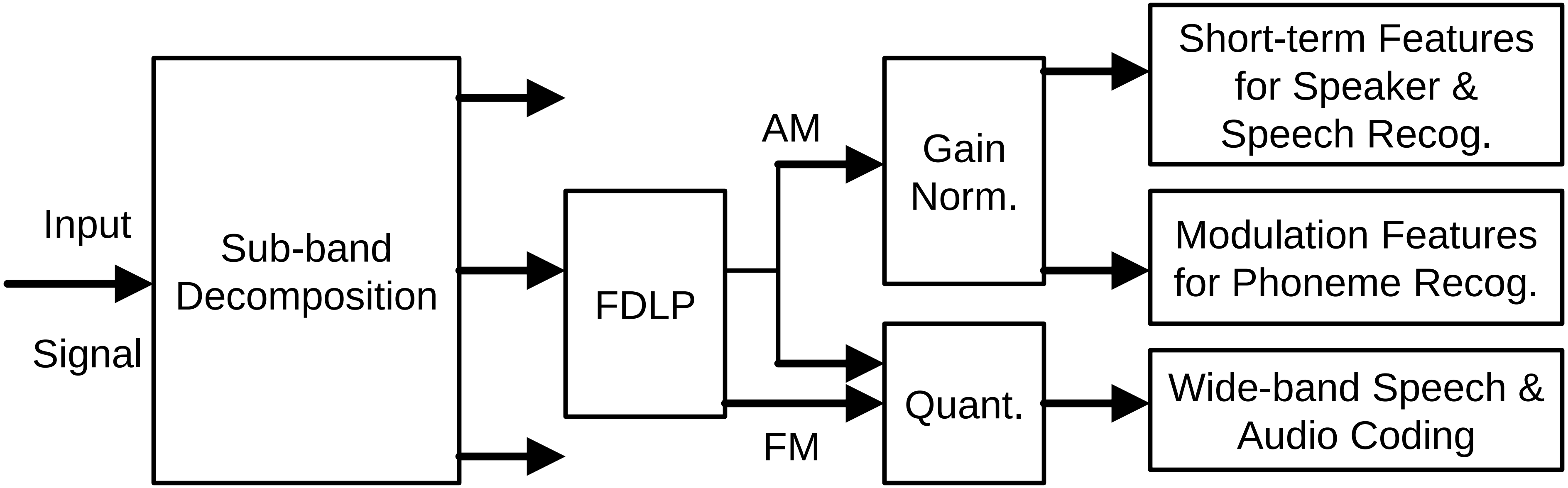
Overview

- Introduction
- AR Model of Hilbert Envelopes
- FDLP and its Properties
- Applications
- Summary

Overview

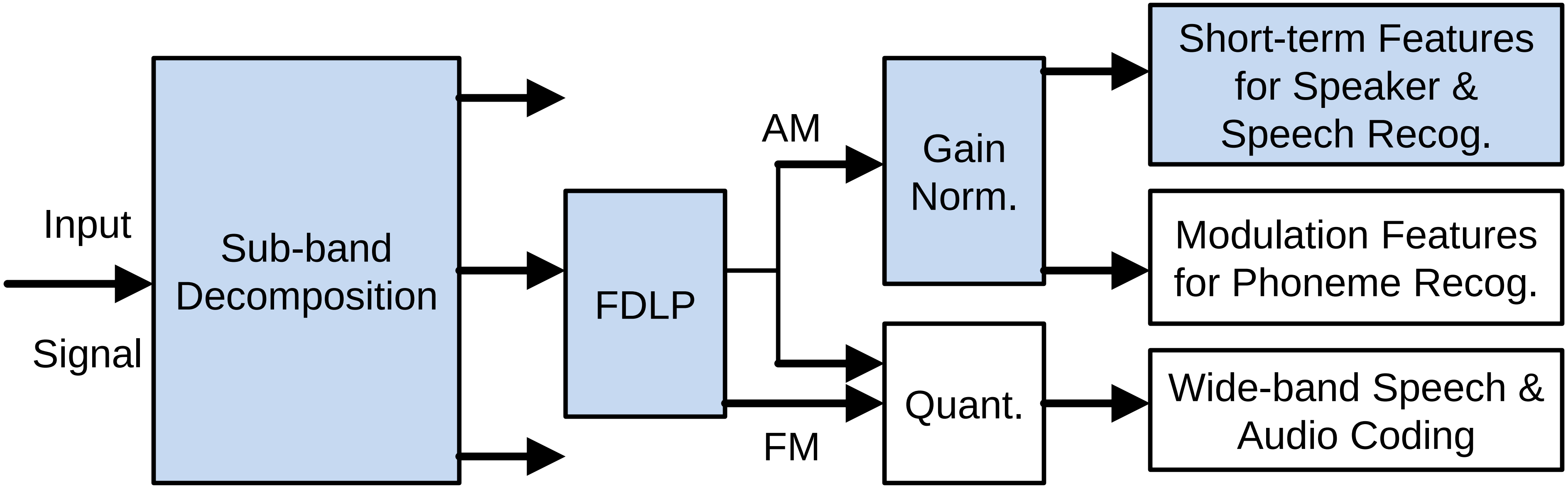
- Introduction
- AR Model of Hilbert Envelopes
- FDLP and its Properties
- Applications
- Summary

Outline of Applications

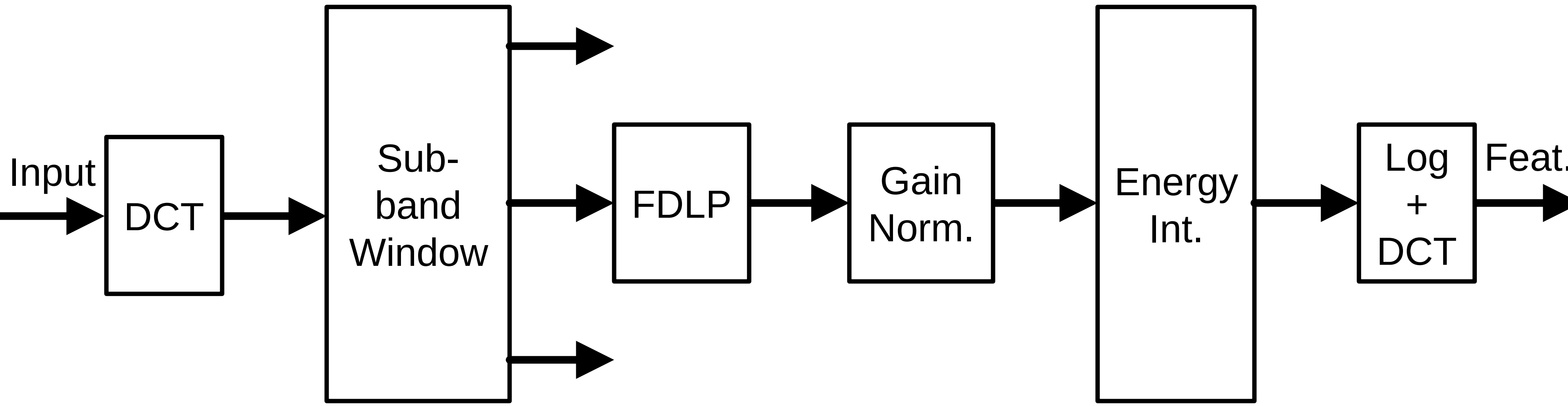


S. Ganapathy, S. Thomas, P. Motlicek and H. Hermansky, "Applications of Signal Analysis Using Autoregressive Models of Amplitude Modulation", *IEEE WASPAA, Oct. 2009.*

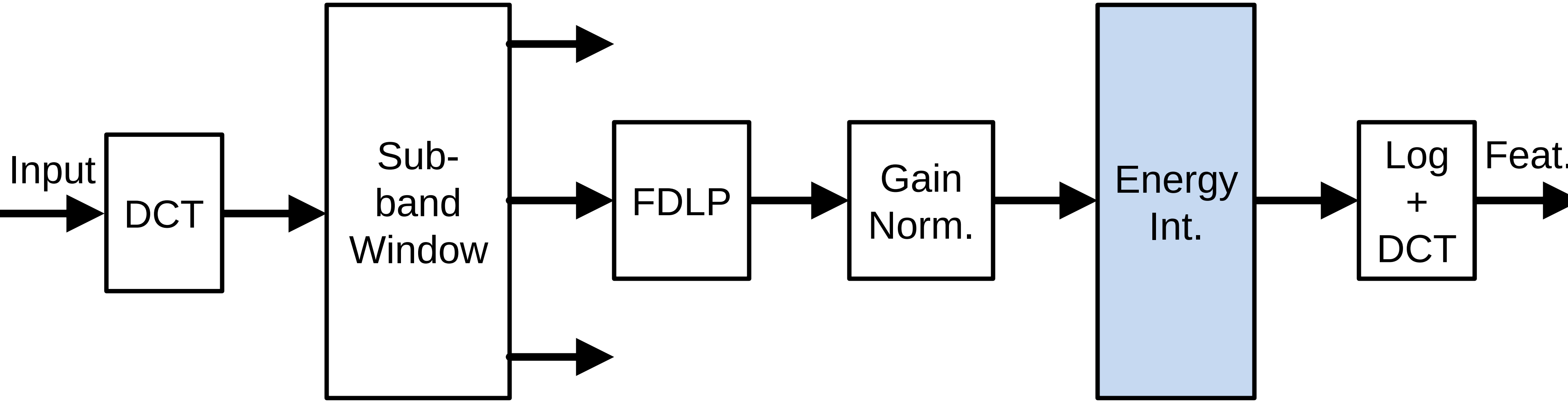
Short-term Features



Short-term Features

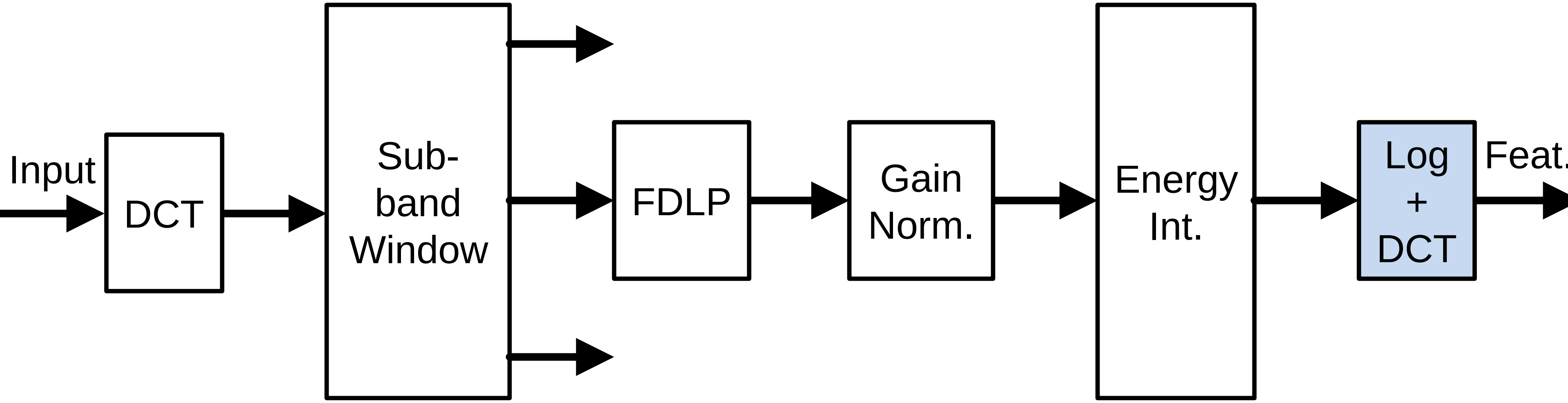


Short-term Features



- Envelopes in each band are integrated along time (25 ms with a shift of 10 ms).
- Integration in frequency axis to convert to mel scale.

Short-term Features



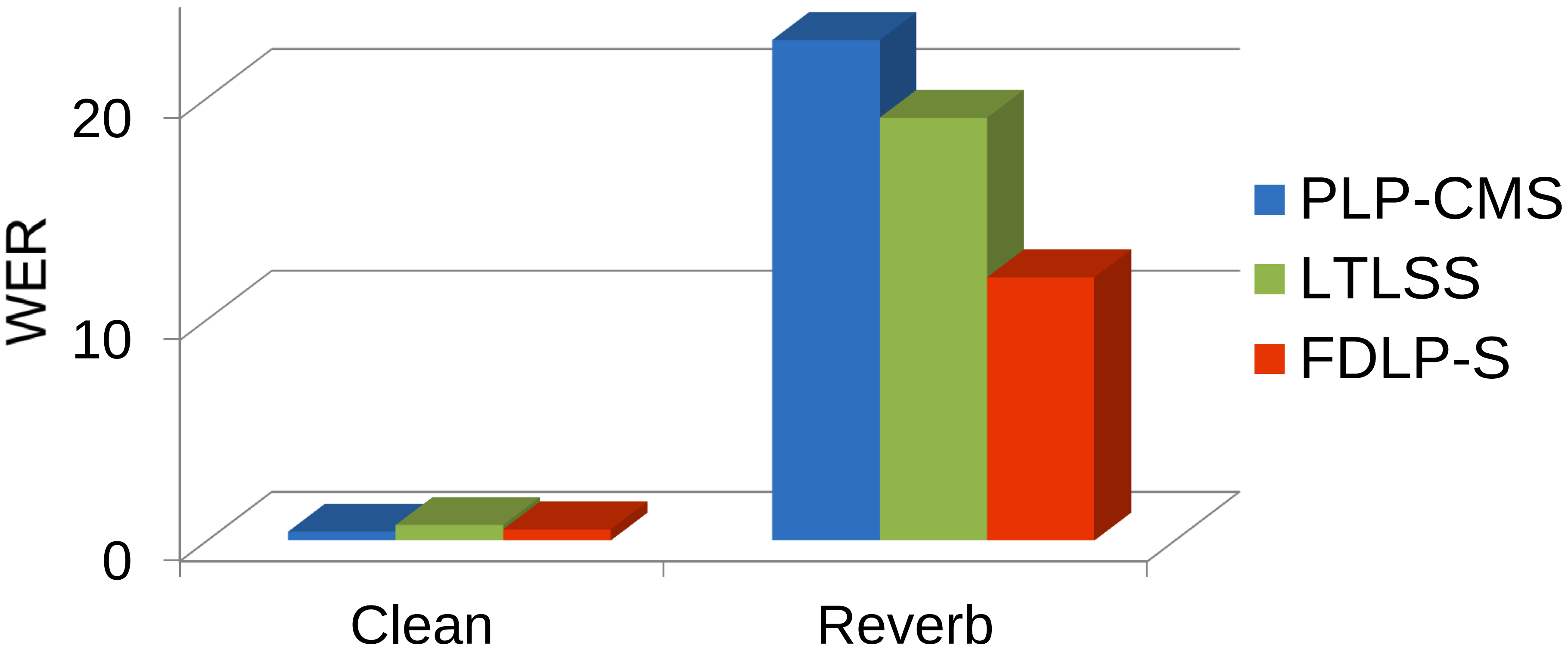
- Sub-band energies are converted to cepstral coefficients by applying log and DCT along frequency axis.
- Delta and acceleration coefficients are appended to obtain 39 dim. feat similar to conventional MFCC feat.

Speech Recognition

- TIDIGITS Database (8 kHz)
  - Clean training data, test data can be clean or naturally reverberated.
- HMM-GMM system
  - Whole-word HMM models trained on clean speech.
  - Performance in terms of word error rate (WER).

- Features
  - PLP features with cepstral mean subtraction (CMS).
  - Long-term log spectral sub. (LTLSS) [Avendano],[Gelbart]
  - FDLP short-term (FDLP-S) features – 39 dim.

Speech Recognition



S. Thomas, S. Ganapathy and H. Hermansky, "Recognition of Reverberant Speech Using FDLP",
IEEE, Signal Proc. Letters, 2008.

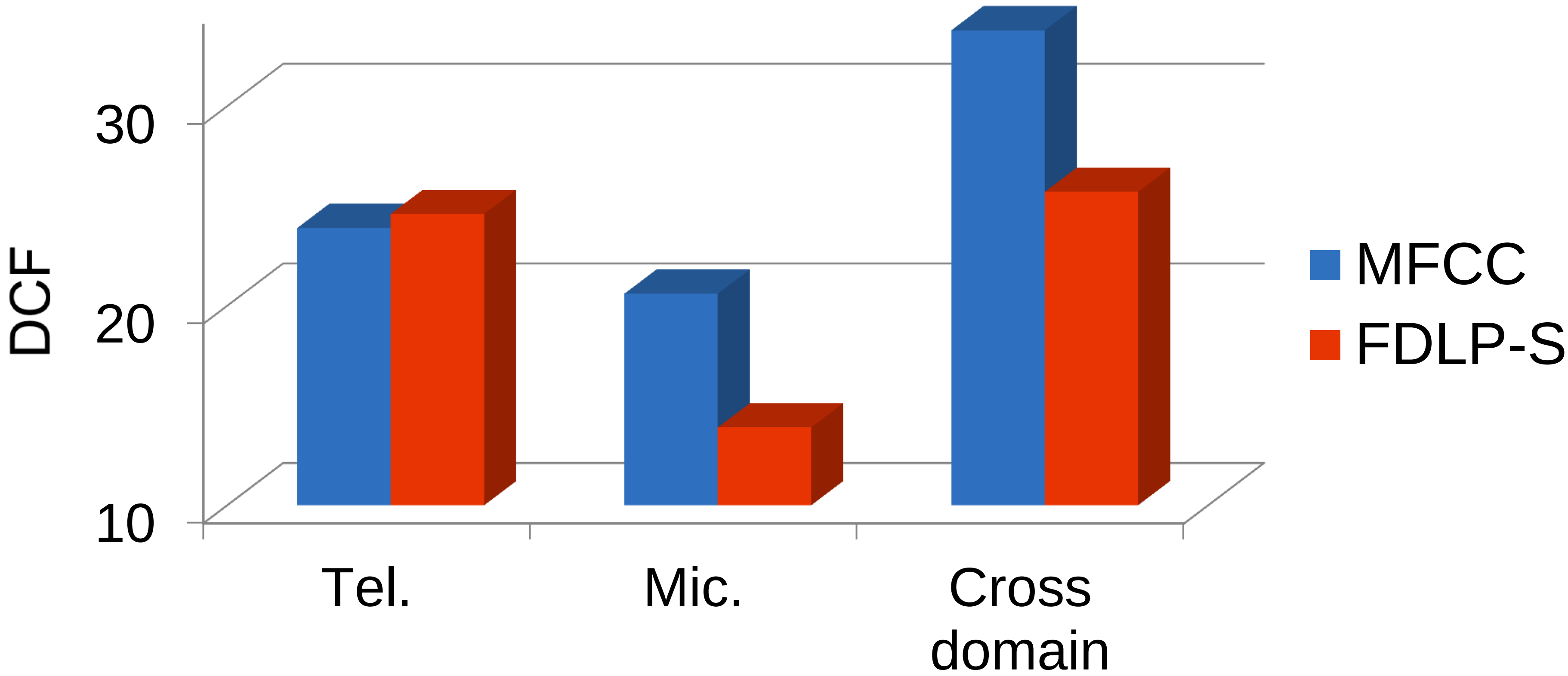
Speaker Verification

- NIST 2008 Speaker recognition evaluation (SRE)
  - Has telephone speech and far-field speech.

- GMM-UBM system
  - Trained on a large set of development speakers.
  - Adapted on the enrollment data from the target speaker.
  - Nuisance attribute projection (NAP) on supervectors.
  - Detection cost function (DCF) $= 0.99\ P_{fa} + 0.1\ P_{miss}$

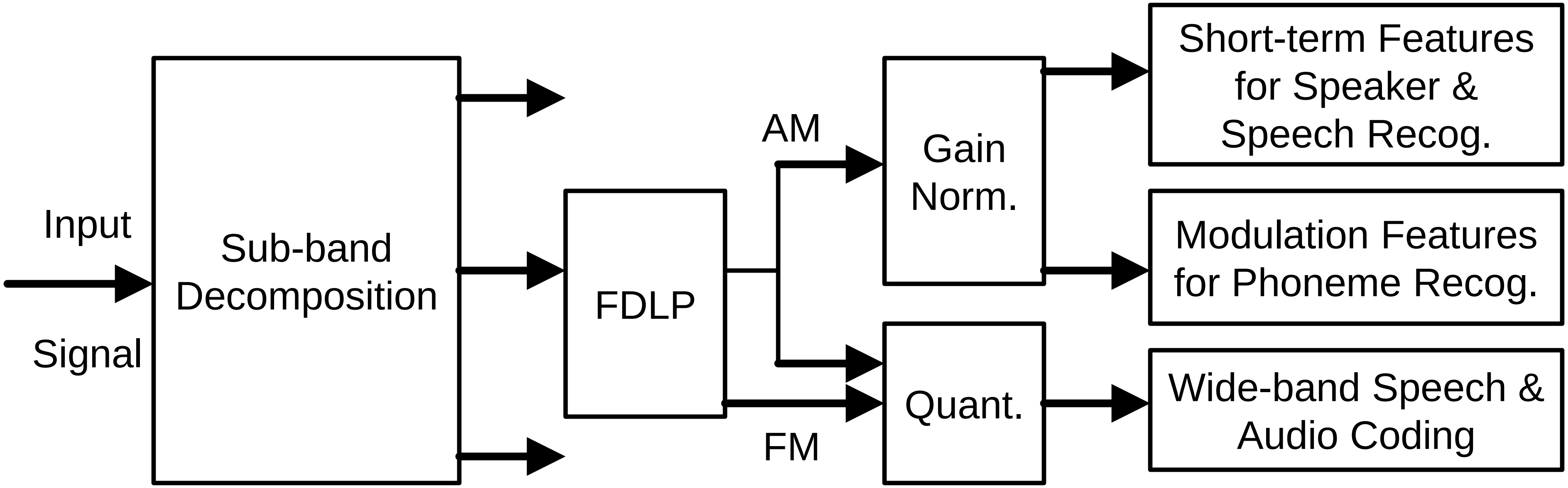
- Features with warping [Pelecanos, 2001].
  - Mel Frequency Cepstral Coefficients (MFCCs)
  - FDLP short-term (FDLP-S) features.

Speaker Verification

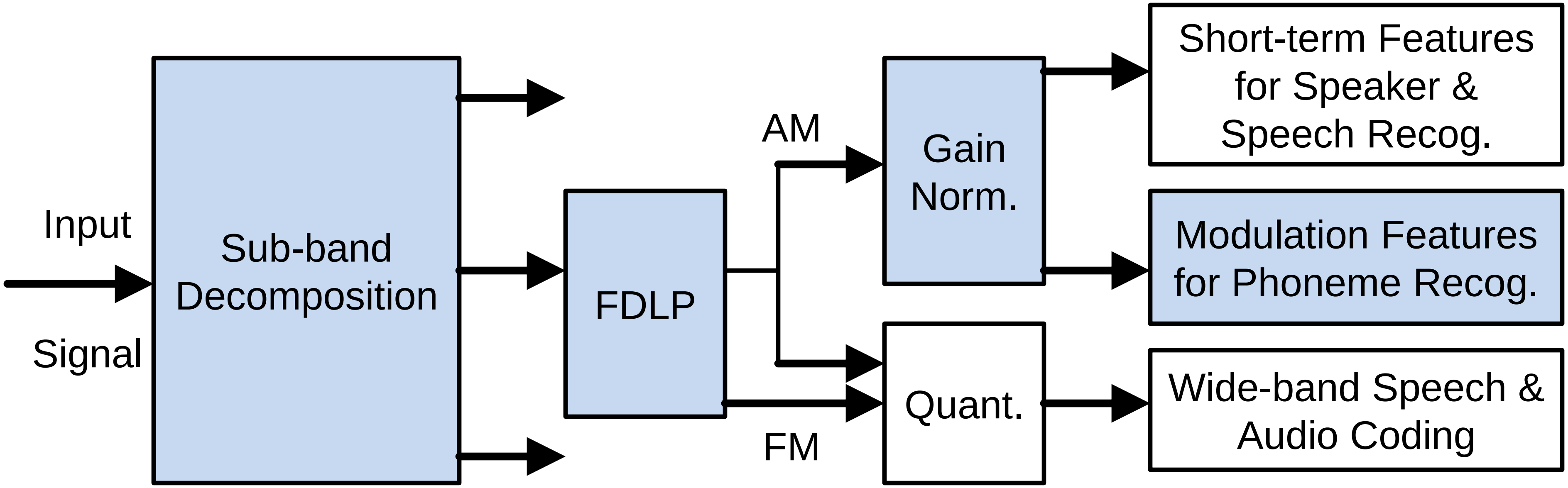


S. Ganapathy, J. Pelecanos and M. Omar, "Feature Normalization for Speaker Verification in Room Reverberation", *ICASSP, 2011.*

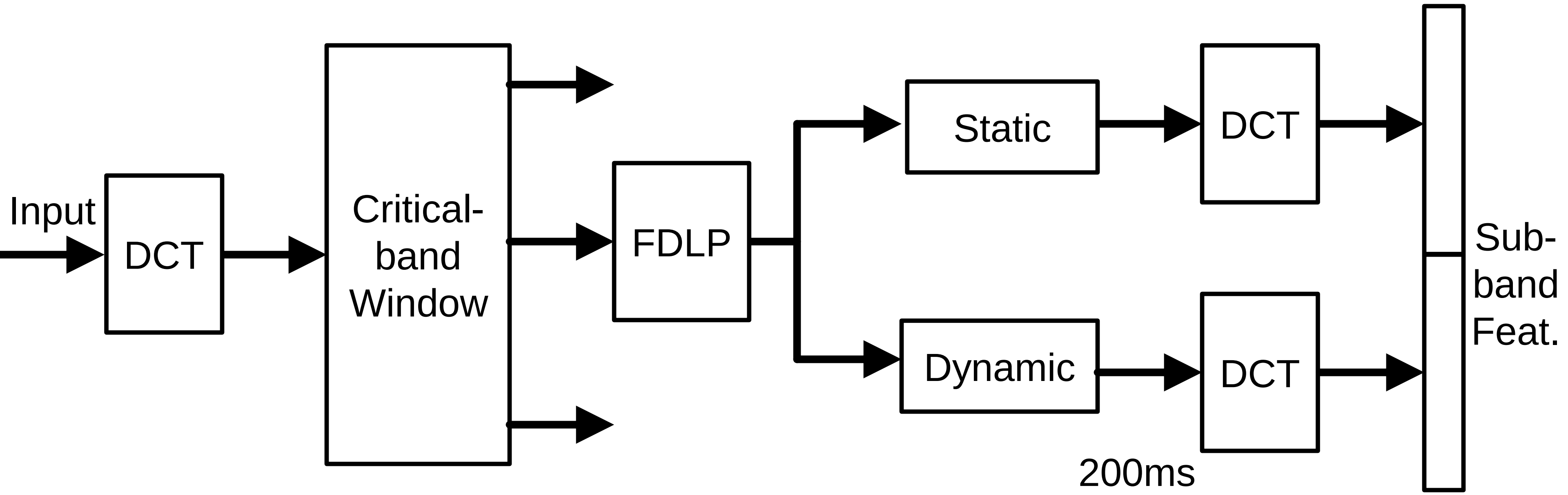
Outline of Applications



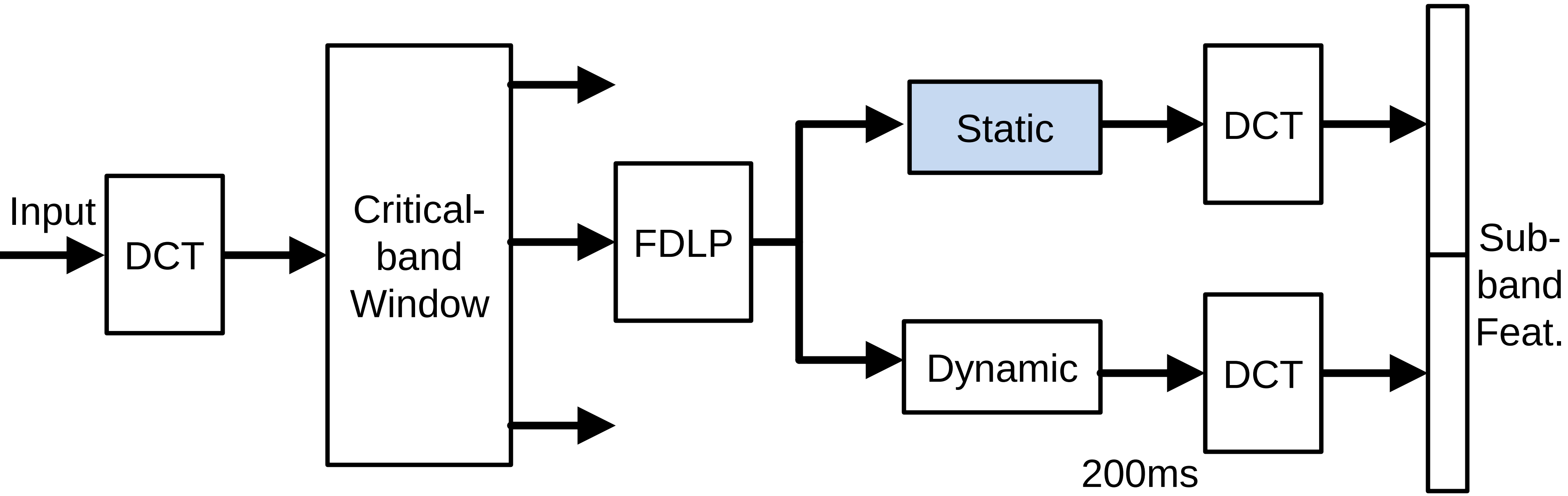
Modulation Features



Modulation Feature Extraction

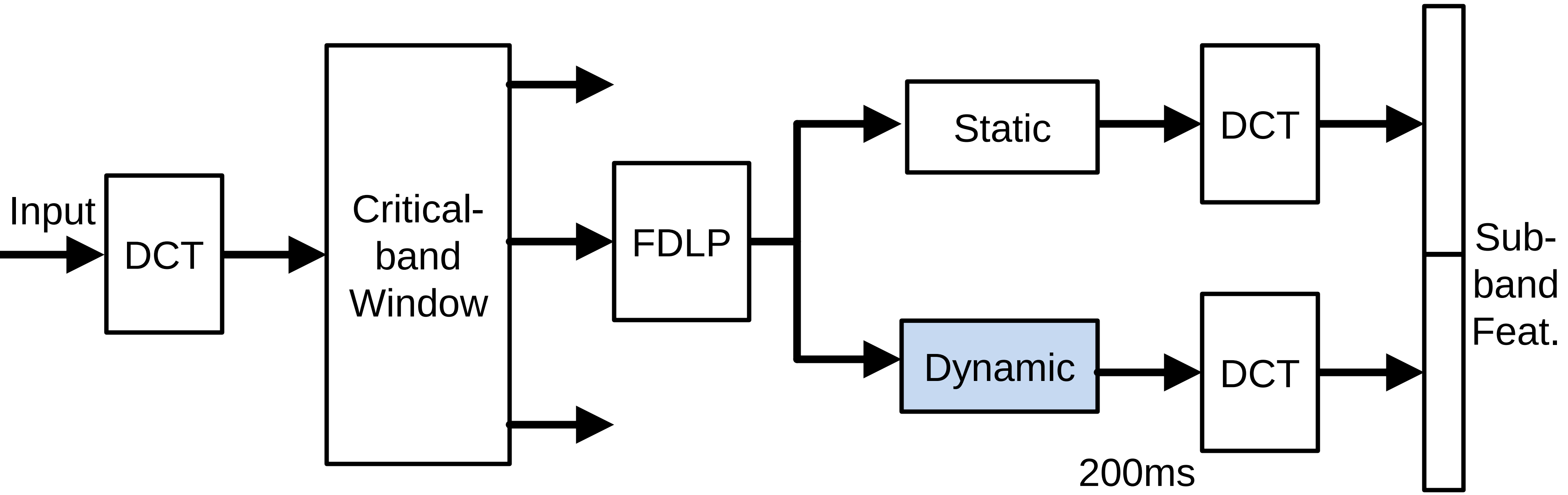


Modulation Feature Extraction



- Static compression is a logarithm – reduce the huge dynamic range in the in the sub-band envelope.

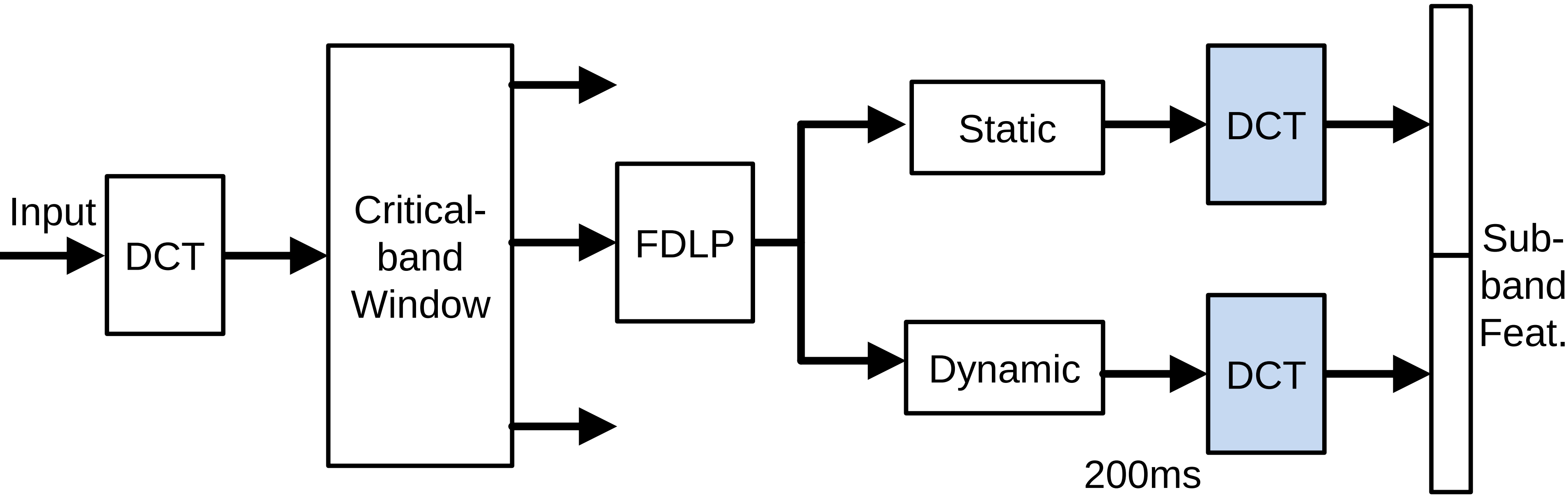
Modulation Feature Extraction



- Dynamic compression is implemented by dynamic compression loops consisting of dividers and low pass filters [Kollmeier, 1999].

.

Modulation Feature Extraction

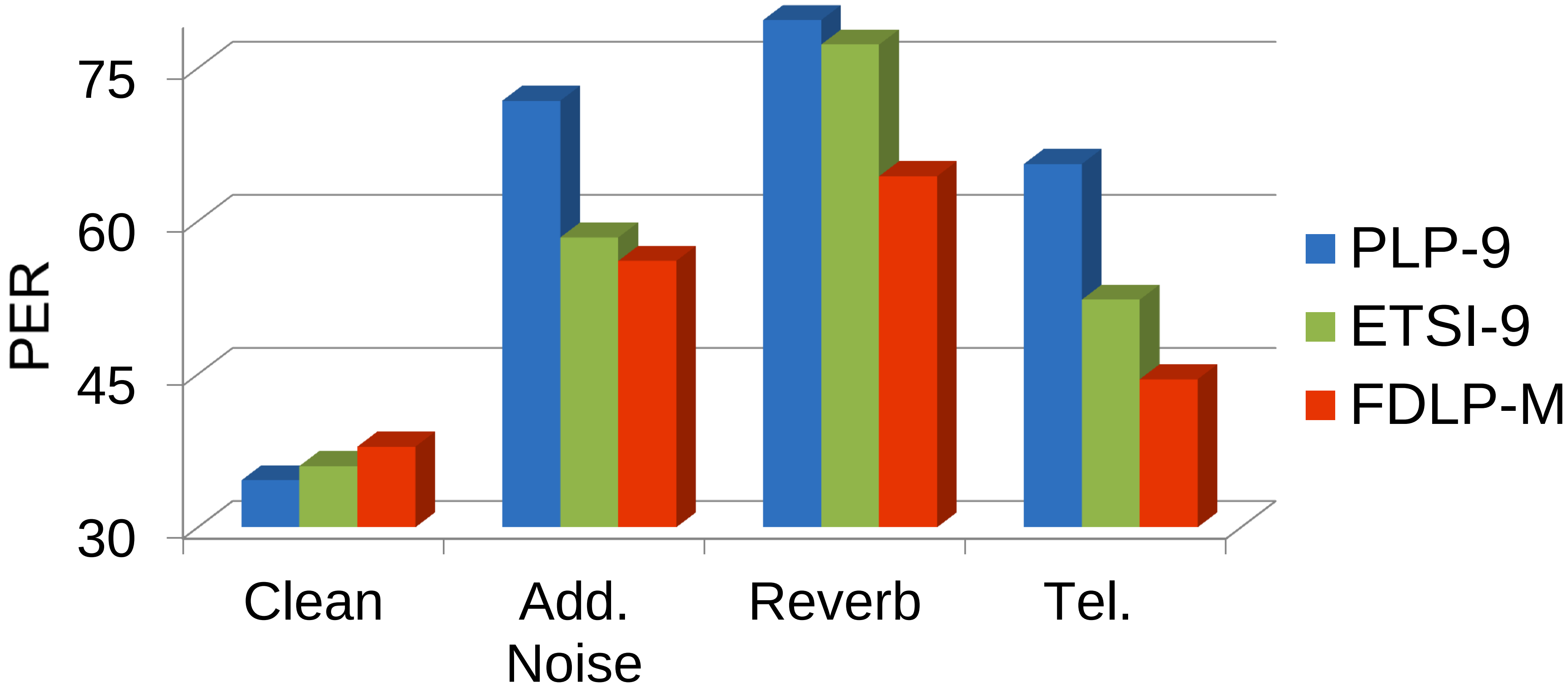


- Compressed sub-band envelopes are DCT transformed to obtain modulation frequency components
- 14 static and dynamic modulation spectra (0-35 Hz) with 17 sub-bands, gets a feature of 476 dim.

Phoneme Recognition

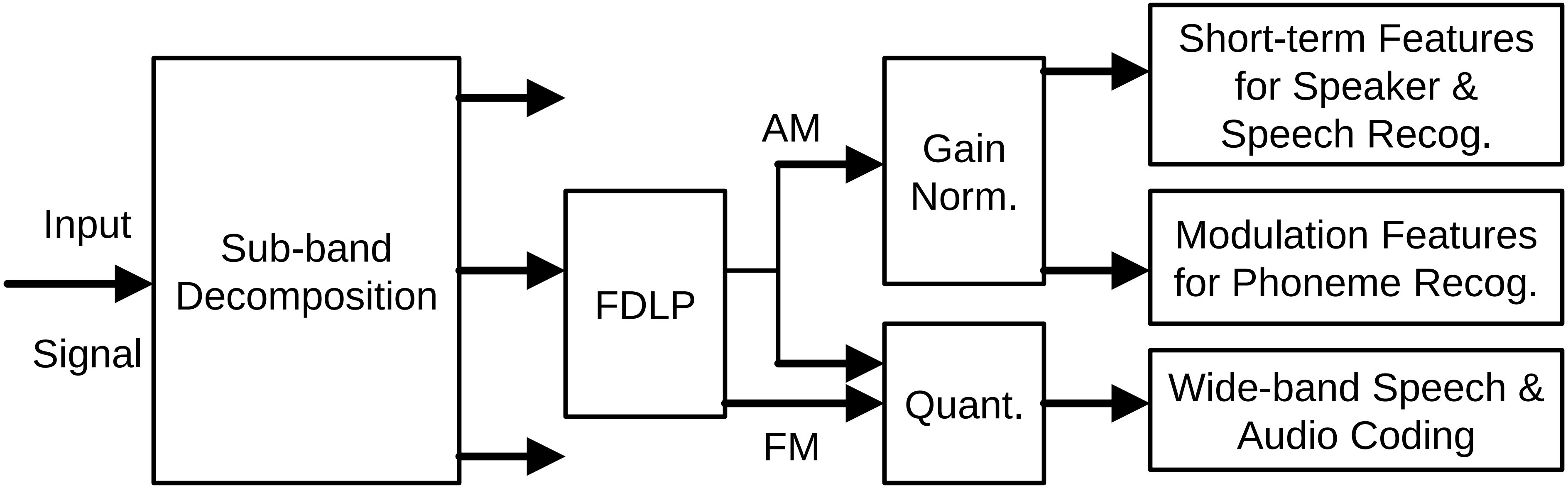
- TIMIT Database (8 kHz)
  - Clean training data, test data can be clean, additive noise, reverberated or telephone channel.
- Multi-layer perceptron (MLP) based system
  - MLPs estimate phoneme posteriors
  - Hidden Markov model (HMM) – MLP hybrid model.
  - Performance in phoneme error rate (PER).
- Features
  - Perceptual linear prediction (PLP) - 9 frame context.
  - Advanced ETSI standard [ETSI,2002] – 9 frame context.
  - FDLP modulation (FDLP-M) features – 476 dim.

Phoneme Recognition

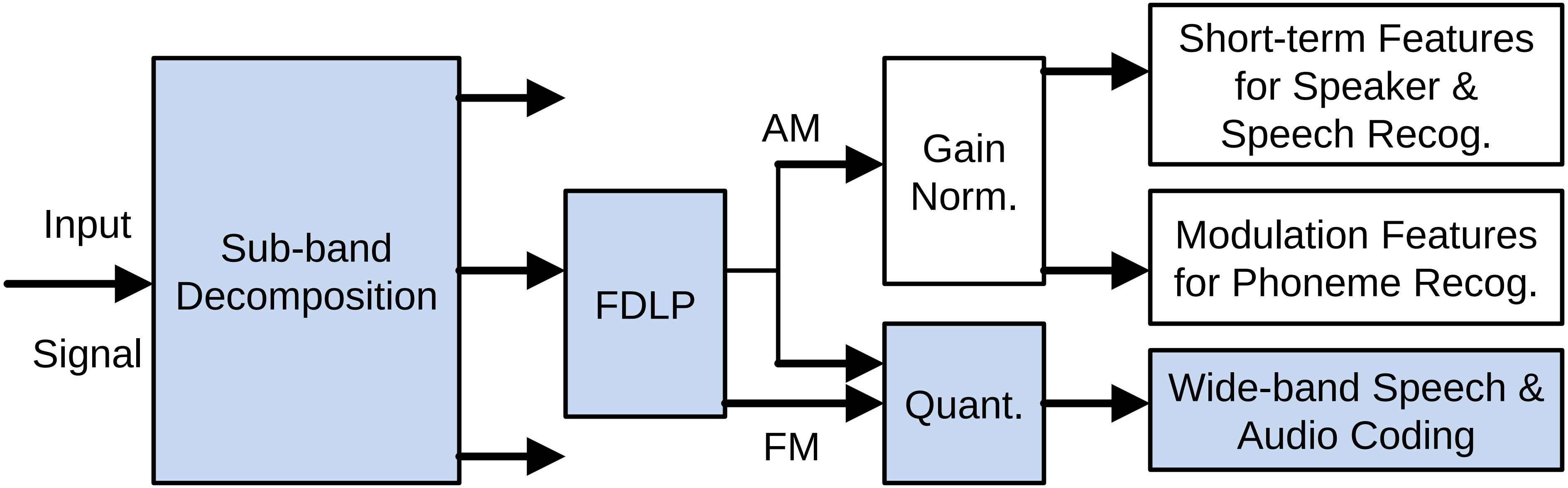


S. Ganapathy, S. Thomas and H. Hermansky, "Temporal Envelope Compensation for Robust Phoneme Recognition Using Modulation Spectrum", *JASA, 2010.*.

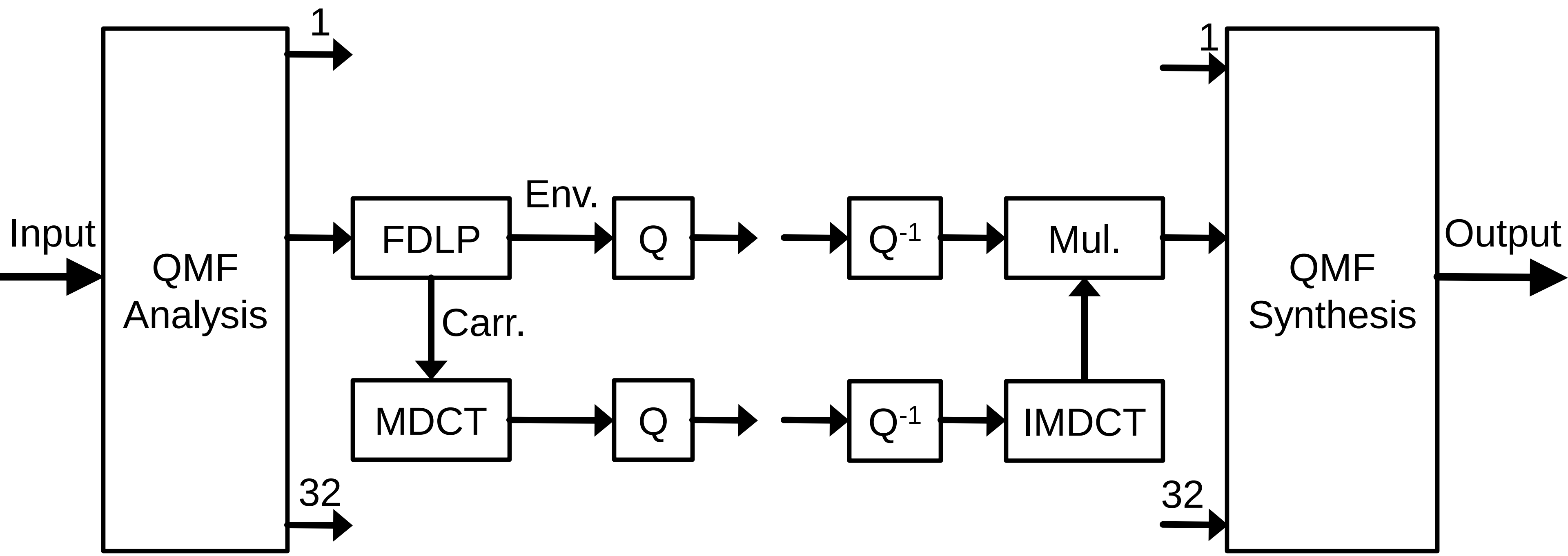
Outline of Applications



Audio Coding

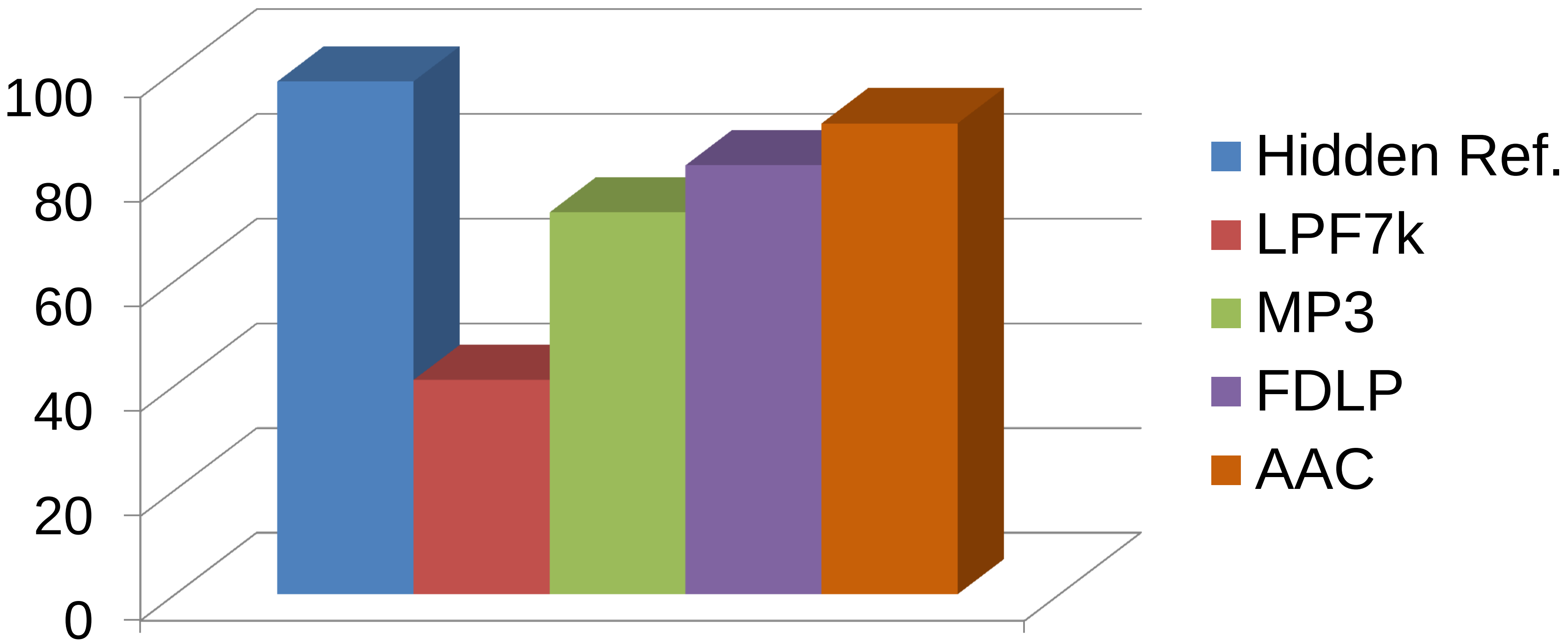


Audio Coding



Sriram Ganapathy, Petr Motlicek and H. Hermansky, “Autoregressive Modeling of Hilbert Envelopes for Wide-band Audio Coding", *AES 124th Convention, Audio Engineering Society, May 2008.*

Subjective Evaluations



S. Ganapathy, P. Motlicek, and H. Hermansky, "AR Models of Amplitude Modulation in Audio Compression", *IEEE Transactions on Audio, Speech and Language Proc., 2010..*

Overview

- Introduction
- AR Model of Hilbert Envelopes
- FDLP and its Properties
- Applications
- Summary

Overview

- Introduction
- AR Model of Hilbert Envelopes
- FDLP and its Properties
- Applications
- Summary

Summary

- Employing AR modeling for estimating amplitude modulations.

- Long-term temporal analysis of signals forms an efficient alternative to conventional short-term spectrum.

- Provides AM-FM decomposition in sub-bands and acts as unified model for speech and audio signals.

Summary

- Employing AR modeling for estimating amplitude modulations.

- Long-term temporal analysis of signals forms an efficient alternative to conventional short-term spectrum.

- Provides AM-FM decomposition in sub-bands and acts as unified model for speech and audio signals.

Summary

- Employing AR modeling for estimating amplitude modulations.
- Long-term temporal analysis of signals forms an efficient alternative to conventional short-term spectrum.
- Provides AM-FM decomposition in sub-bands and acts as unified model for speech and audio signals.

Our Contributions

- Simple mathematical analysis for AR model of Hilbert envelopes.
- Investigating the resolution properties of FDLP.
- Gain normalization of FDLP Envelopes

Our Contributions

- Simple mathematical analysis for AR model of Hilbert envelopes.
- Investigating the resolution properties of FDLP.
- Gain normalization of FDLP Envelopes

Our Contributions

- Simple mathematical analysis for AR model of Hilbert envelopes.

- Investigating the resolution properties of FDLP.

- Gain normalization of FDLP Envelopes

Our Contributions

- Short-term feature extraction using FDLP –Improvements in reverb speech recog.
- Modulation feature extraction – Phoneme recognition in noisy speech.
- Speech and audio codec development using AM-FM signals from FDLP.

Our Contributions

- Short-term feature extraction using FDLP –Improvements in reverb speech recog.
- Modulation feature extraction – Phoneme recognition in noisy speech.
- Speech and audio codec development using AM-FM signals from FDLP.

Our Contributions

- Short-term feature extraction using FDLP –Improvements in reverb speech recog.
- Modulation feature extraction – Phoneme recognition in noisy speech.
- Speech and audio codec development using AM-FM signals from FDLP.

Publications

Journals

S. Ganapathy, S. Thomas and H. Hermansky, "Temporal envelope compensation for robust phoneme recognition using modulation spectrum ", Journal of Acoustical Society of America, Dec. 2010.

S. Ganapathy, P. Motlicek and H. Hermansky, "Autoregressive Models Of Amplitude Modulations In Audio Compression", IEEE Transactions on Audio, Speech and Language Processing, Aug. 2010.

P. Motlicek, **S. Ganapathy**, H. Hermansky and H. Garudadri,"Wide-Band Audio Coding based on Frequency Domain Linear Prediction", EURASIP Journal on Audio, Speech, and Music Processing, 2010.

S. Ganapathy, S. Thomas and H. Hermansky, "Modulation Frequency Features For Phoneme Recognition In Noisy Speech", Journal of Acoustical Society of America - Express Letters, Jan 2009.

S. Thomas, **S. Ganapathy** and H. Hermansky, "Recognition Of Reverberant Speech Using Frequency Domain Linear Prediction", IEEE Signal Processing Letters, Dec 2008.

Patents

Temporal Masking in Audio Coding Based on Spectral Dynamics in Frequency Sub-bands

"Spectral Noise Shaping in Audio Coding Based on Spectral Dynamics in Frequency Sub-bands

Publications

Selected Conferences

S. Ganapathy, P. Rajan and H. Hermansky, "Multi-layer Perceptron Based Speech Activity Detection for Speaker Verification", IEEE WASPAA, Oct. 2011.

S. Ganapathy, J. Pelecanos and M. Omar, "Feature Normalization for Speaker Verification in Room Reverberation", ICASSP, May 2011.

S. Ganapathy, S. Thomas and H. Hermansky, "Robust Spectro-Temporal Features Based on Autoregressive Models of Hilbert Envelopes", ICASSP, March 2010.

S. Ganapathy, S. Thomas and H. Hermansky, "Comparison of Modulation Features For Phoneme Recognition", ICASSP, March 2010.

S. Ganapathy, S. Thomas, and H. Hermansky, "Temporal Envelope Subtraction for Robust Speech Recognition Using Modulation Spectrum", IEEE ASRU, 2009.

S. Ganapathy, S. Thomas, P. Motlicek and H. Hermansky, "Applications of Signal Analysis Using Autoregressive Models for Amplitude Modulation", IEEE WASPAA 2009.

S. Ganapathy, S. Thomas and H. Hermansky, "Static and Dynamic Modulation Spectrum for Speech Recognition", Proc. of INTERSPEECH, Brighton, UK, Sept. 2009.

S. Ganapathy, P. Motlicek, H. Hermansky and H. Garudadri, "Autoregressive Modelling of Hilbert Envelopes for Wide-band Audio Coding", AES 124th Convention, AES.

S. Ganapathy, P. Motlicek, H. Hermansky and H. Garudadri, ""Temporal Masking for Bit-rate Reduction in Audio Codec Based on Frequency Domain Linear Prediction", ICASSP, April 2008.

Acknowledgements

- Lab Buddies – Samuel Thomas, Sivaram Garimella, Padmanbhan Rajan, Harish Mallidi, Vijay Peddinti, Thomas Janu, Aren Jansen.

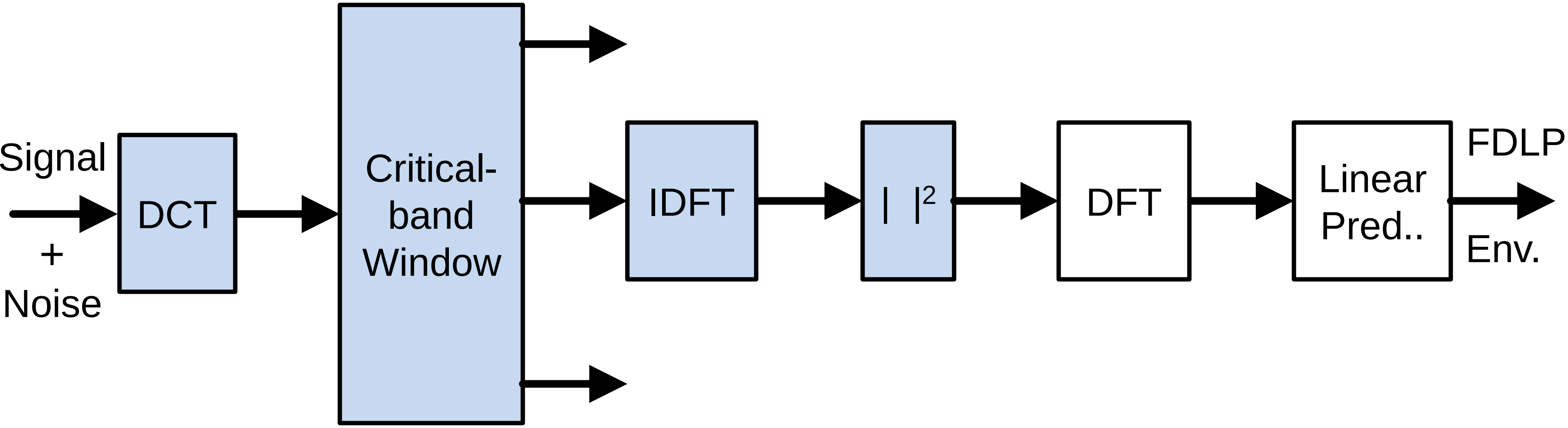
- Idiap personnel – Petr Motlicek, Joel Pinto, Mathew Doss.

- IBM personnel – Jason Pelecanos, Mohamed Omar

- Others – Xinhui Zhou, Daniel Romero, Marios Athineos, David Gelbart, Harinath Garudadri.

Thank You

Noise Compensation in FDLP

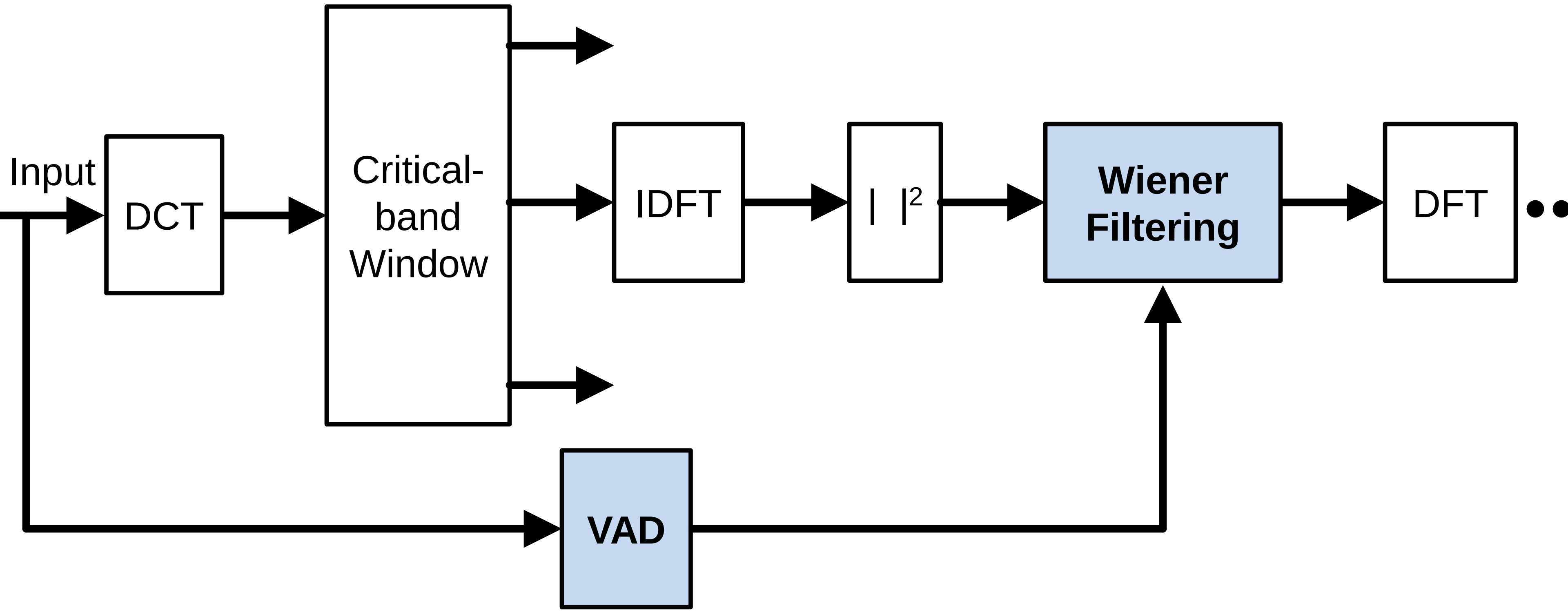


- When speech is corrupted with additive noise,

$$y[n] = x[n] + s[n]$$

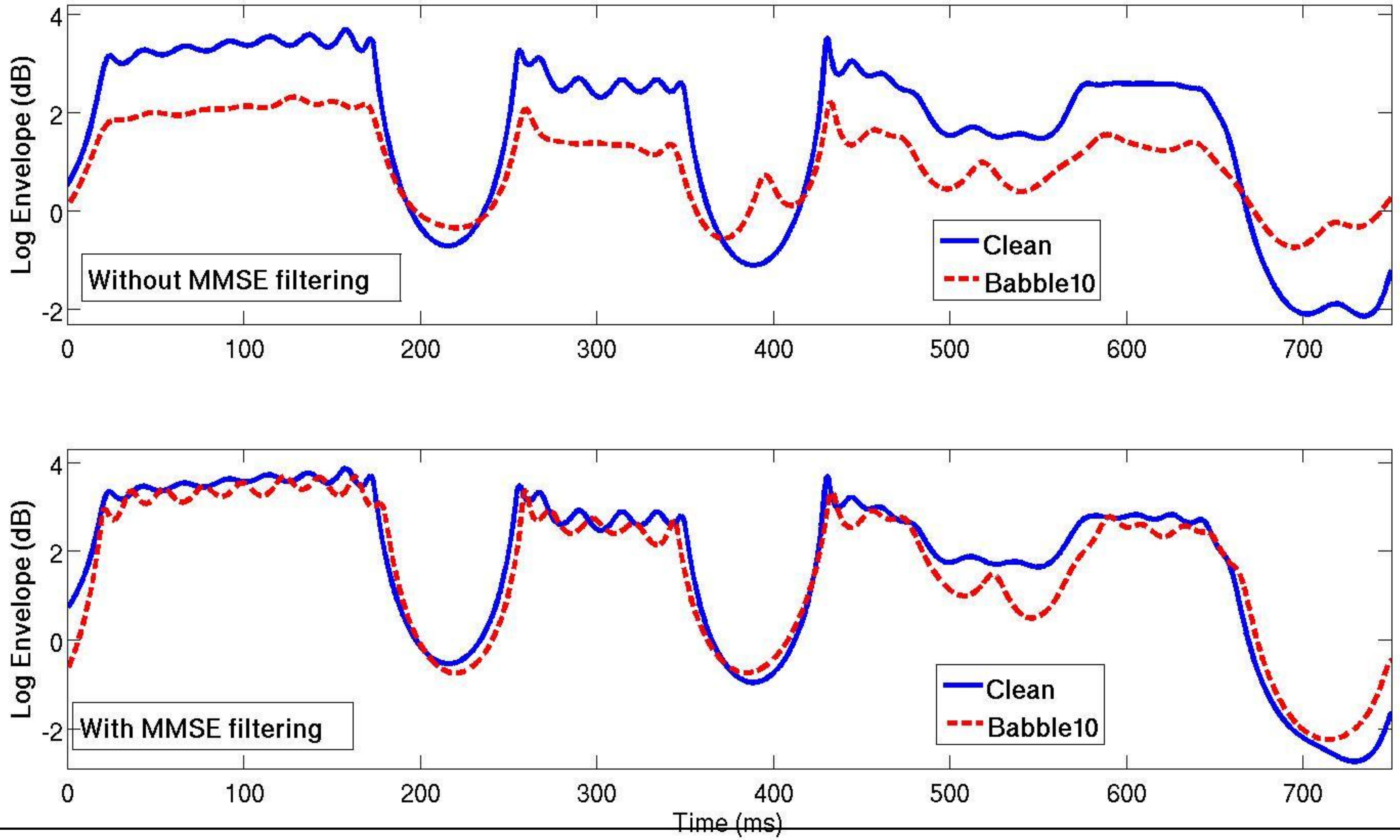
- The noise component is additive in the non-parametric Hilbert envelope domain (assuming the signal and noise are uncorrelated).

Noise Compensation in FDLP



- Voice activity detector (VAD) provides information about the non-speech regions which are used for estimating the temporal envelope of the noise.
- Noise subtraction tries to subtract the estimate the noise envelope from the noisy speech envelope.

Noise Compensation in FDLP



S. Ganapathy, S. Thomas, and H. Hermansky, "Temporal Envelope Subtraction for Robust Speech Recognition using Modulation Spectrum", *IEEE ASRU, 2009.*

Dealing with Convolutive Distortions

- Cepstral mean subtraction (CMS), long-term log spectral subtraction (LTLSS) & gain normalization
  - CMS assumes distortion in neighboring frames to be similar – suppresses short-term artifacts.
  - Long-term subtraction deals with reverberation assuming over the same response over a window of long-term frames [Gelbart, 2002].
  - Gain normalization deals with short and long term distortions within a single long-term frame.

Dealing with Convolutive Distortions

- Cepstral mean subtraction (CMS), long-term log spectral subtraction (LTLSS) & gain normalization
  - CMS assumes distortion in neighboring frames to be similar – suppresses short-term artifacts.
  - Long-term subtraction deals with reverberation assuming over the same response over a window of long-term frames [Gelbart, 2002].
  - Gain normalization deals with short and long term distortions within a single long-term frame.

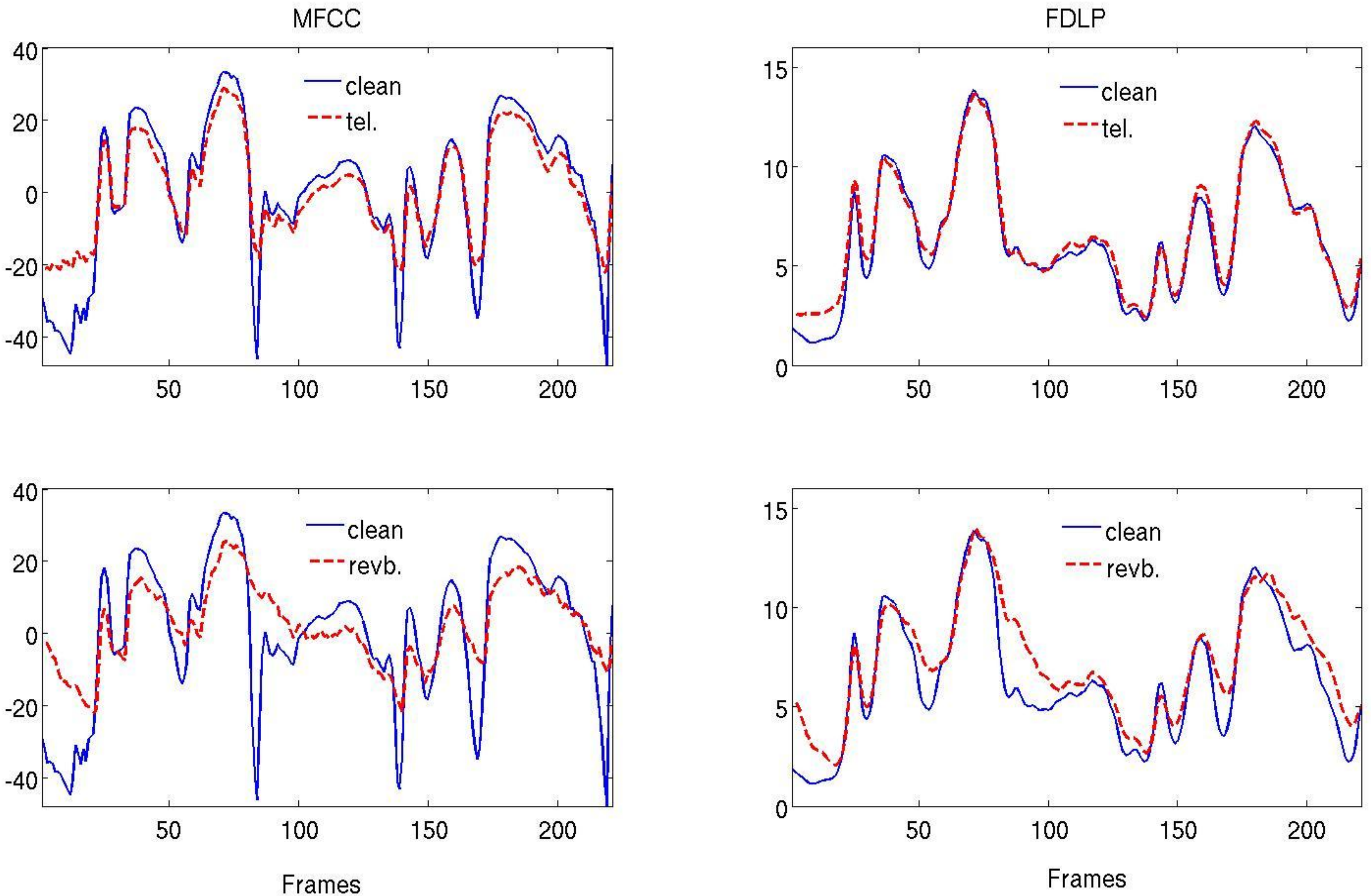
Dealing with Convolutive Distortions

- Cepstral mean subtraction (CMS), long-term log spectral subtraction (LTLSS) & gain normalization
  - CMS assumes distortion in neighboring frames to be similar – suppresses short-term artifacts.
  - Long-term subtraction deals with reverberation assuming over the same response over a window of long-term frames [Gelbart, 2002].
  - Gain normalization deals with short and long term distortions within a single long-term frame.

Dealing with Convolutive Distortions

- Cepstral mean subtraction (CMS), long-term log spectral subtraction (LTLSS) & gain normalization
  - CMS assumes distortion in neighboring frames to be similar – suppresses short-term artifacts.
  - Long-term subtraction deals with reverberation assuming over the same response over a window of long-term frames [Gelbart, 2002].
  - Gain normalization deals with short and long term distortions within a single long-term frame.

Feature Comparison



Evidences

- Physiological evidences -
  - Spectro-temporal receptive fields [Shamma et.al. 2001]

- Psycho-physical evidences -
  - Perceptual importance of modulation frequencies [Drullman et al. 1994].
  - Syllable recognition from temporal modulations with minimal spectral cues [Shannon et al., 1995].

Evidences

- Physiological evidences -
  - Spectro-temporal receptive fields [Shamma et.al. 2001].

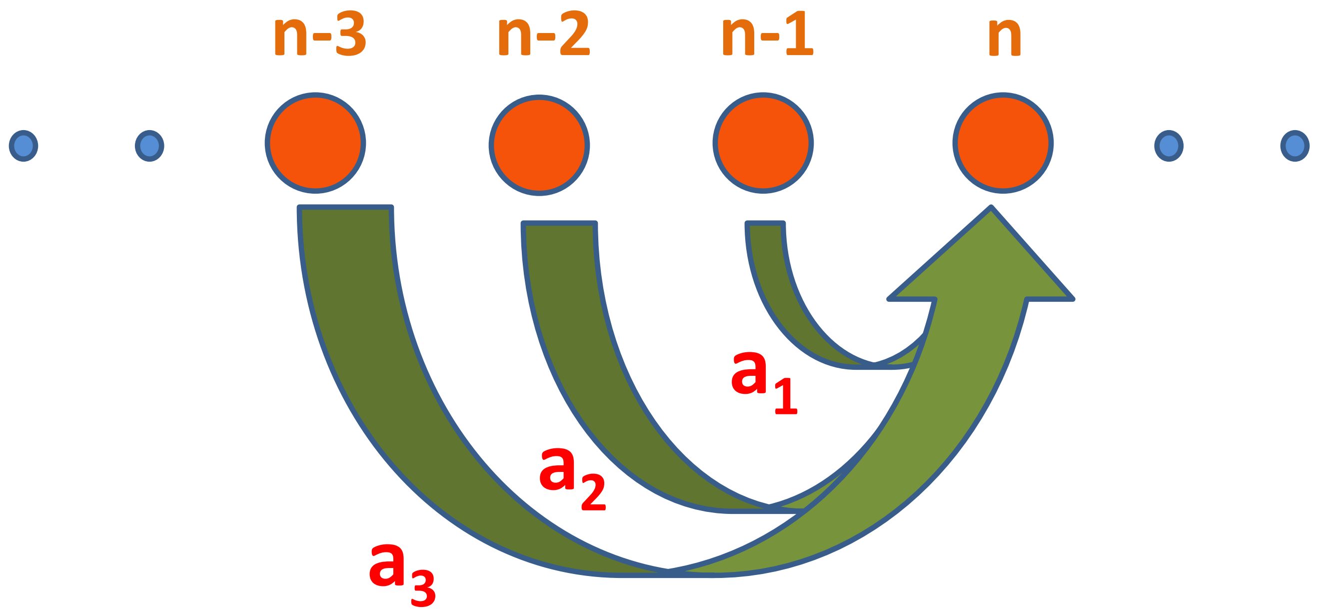
- Psycho-physical evidences -
  - Perceptual importance of modulation frequencies [Drullman et al. 1994].
  - Syllable recognition from temporal modulations with minimal spectral cues [Shannon et al., 1995].

Applications

- Modulation spectra has been used in the past
    - Speech intelligibility [Houtgast et al, 1980].
    - RASTA processing [Hermansky et al, 1994].
    - Speech recognition [Kingsbury et al, 1998].
    - AM-FM decomposition [Kumaresan et al, 1999].
    - Sound texture modeling [Athineos et al, 2003].
    - Sound source separation [King et al, 2010].

Linear Prediction – Time Domain

- Current sample expressed as a linear combination of past samples



Linear Prediction – Time Domain

- Current sample expressed as a linear combination of past samples

$$x[n] = \sum_{k=1}^{p} a_k x[n-k] + \boldsymbol{e}[n] \quad \forall\, n = 0 \ldots N-1$$

- Model parameters are solved by minimizing the residual sum of squares.

$$E_p = \sum_{n=0}^{N-1} |\boldsymbol{e}[n]|^2$$

AR model of Power Spectrum

Filter interpretation [Makhoul, 1975]

$$\boldsymbol{e}[n] = x[n] - \sum_{i=1}^{p} a_i x[n-i] = x[n] * d[n]$$

$$d = [1 \;\; -a_1 \;\; -a_2 \ldots -a_p]$$

$$\boldsymbol{\mathcal{E}}(\omega) = \sum_{n=0}^{N-1} \boldsymbol{e}[n] e^{-j\omega n} = X(\omega) D(\omega)$$

From Parseval's theorem

$$E_p = \sum_{n=0}^{N-1} |\boldsymbol{e}[n]|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\boldsymbol{\mathcal{E}}(\omega)|^2 \, d\omega$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(\omega)|^2 \, |D(\omega)|^2 \, d\omega$$

AR model of Power Spectrum

By definition,

$$|D(\omega)|^2 = |1 - \sum_{i=1}^{p} a_i e^{-ji\omega}|^2$$

Let,

$$P_x(\omega) = |X(\omega)|^2, \quad H(\omega) = \frac{1}{D(\omega)}$$

Thus, parameters $\{a_i\}$ are solved by minimizing

$$E_p = \frac{1}{2\pi}\int_{-\pi}^{\pi} |X(\omega)|^2 \, |D(\omega)|^2 \; d\omega = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{P_x(\omega)}{|H(\omega)|2} \, d\omega$$

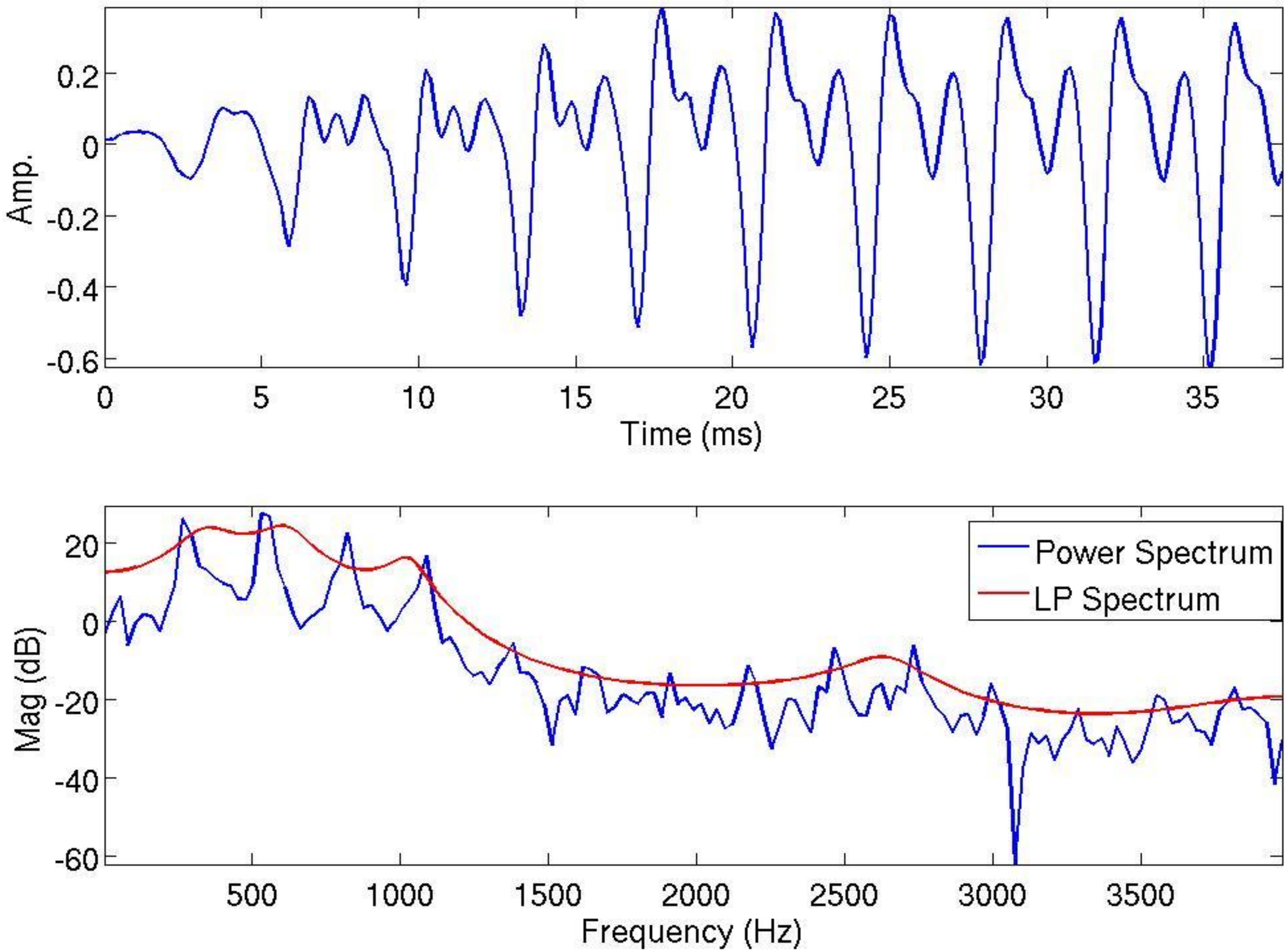
AR model of Power Spectrum

- Solution of the linear prediction yields an all-pole model of the power spectrum

$$\widehat{P_x}[\omega] = Ep\,|H(\omega)|2 = \frac{G}{|1 - \sum_{i=1}^{p} a_i e^{-ji\omega}|^2}$$

- Numerator G denotes the gain of AR model (equal to minimum residual sum of squares).

AR model of power spectrum



Hilbert Envelope - Definition

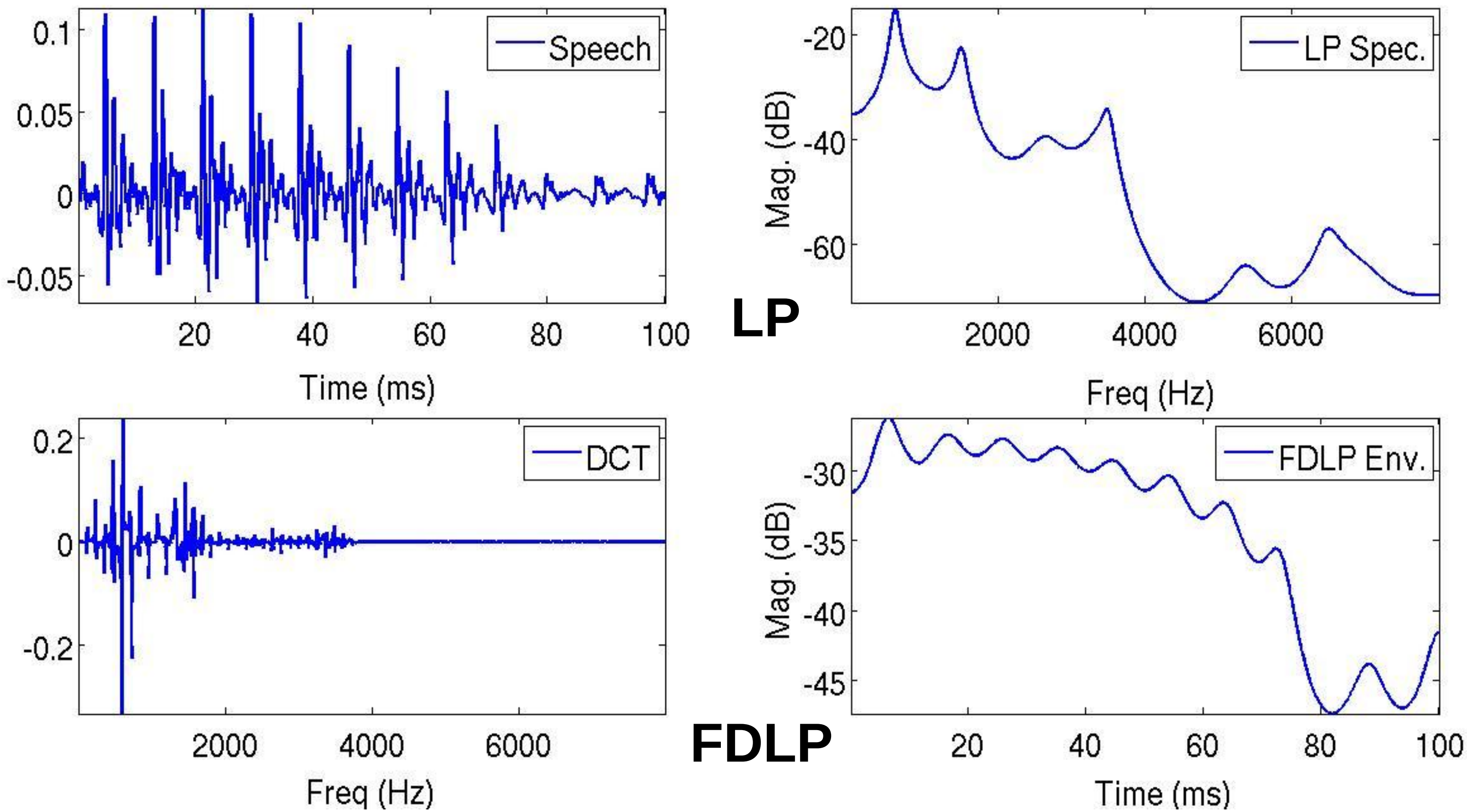
- Analytic signal is the sum of the signal and its quadrature component.

$$x_a[n] = x[n] + j\mathscr{H}(x[n])$$

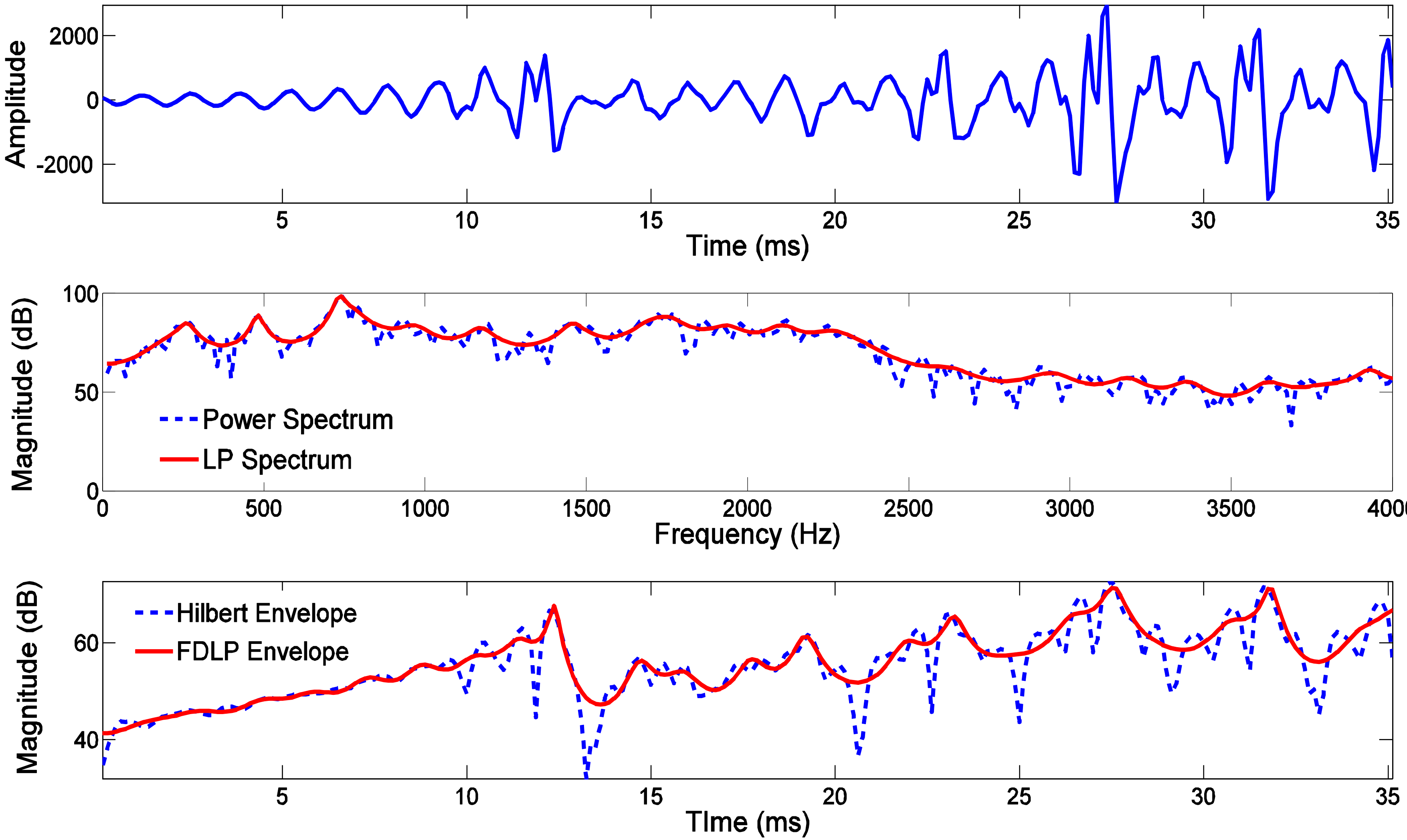
where $\mathscr{H}$ denotes the Hilbert transform.

- Hilbert envelope is the squared magnitude of the analytic signal.

Duality

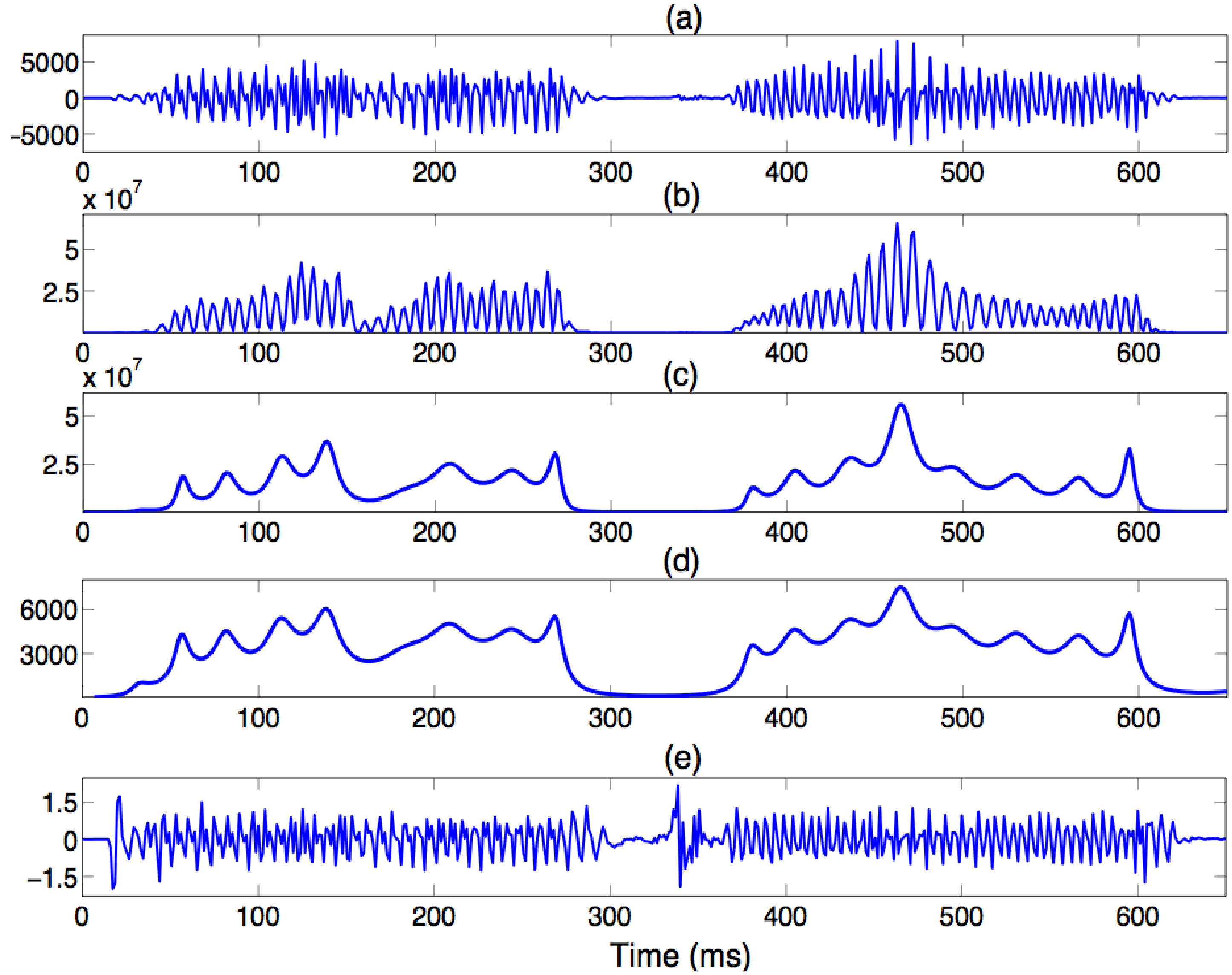

Speech
Time (ms)
LP Spec.
Mag. (dB)
Freq (Hz)
LP
DCT
Freq (Hz)
FDLP Env.
Mag. (dB)
Time (ms)
FDLP

LP in Time and Frequency

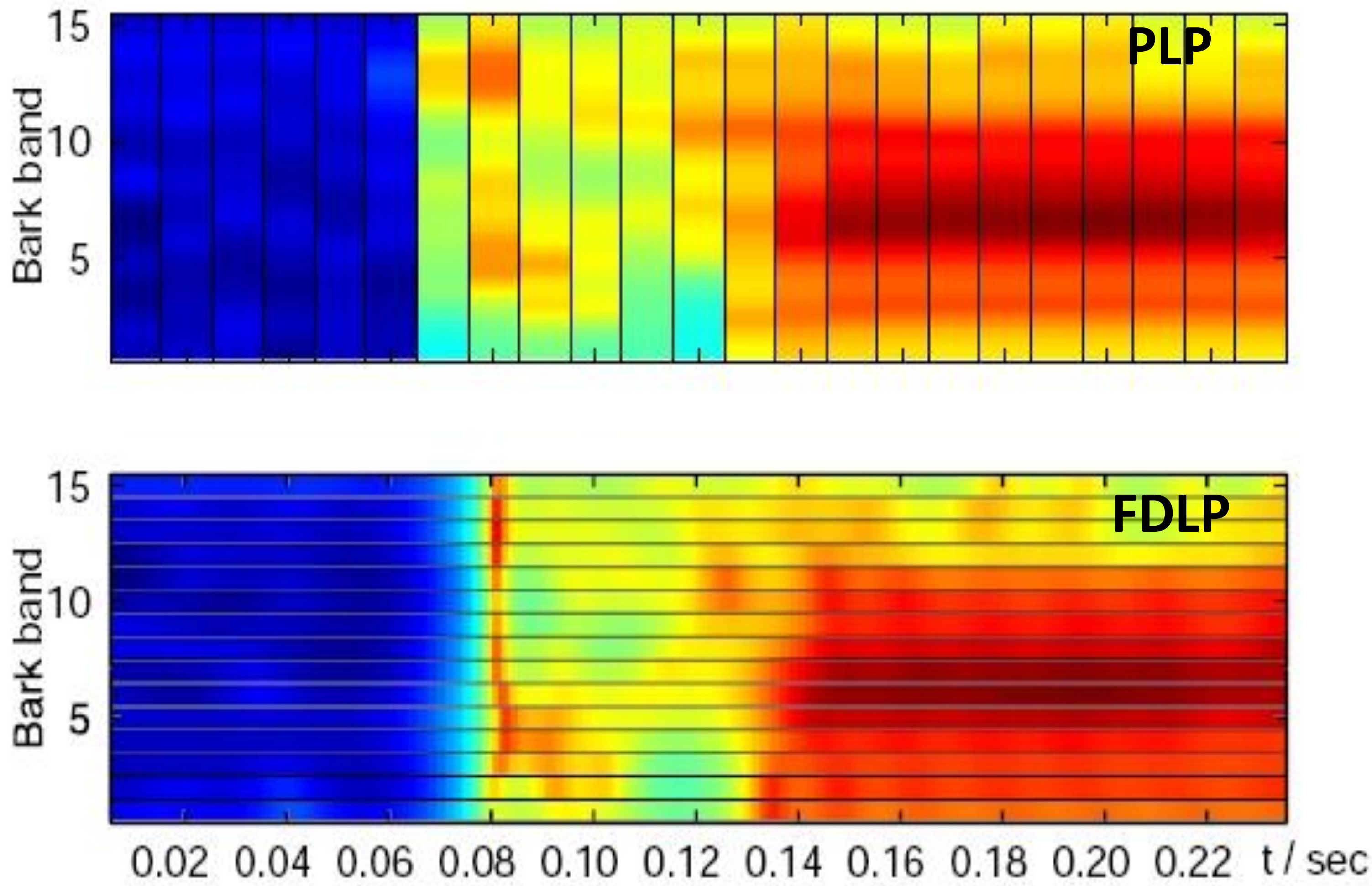


AM-FM Decomposition

a. Signal
b. Hilb. Env.
c. FDLP Env.
d. AM comp.
e. FM comp.

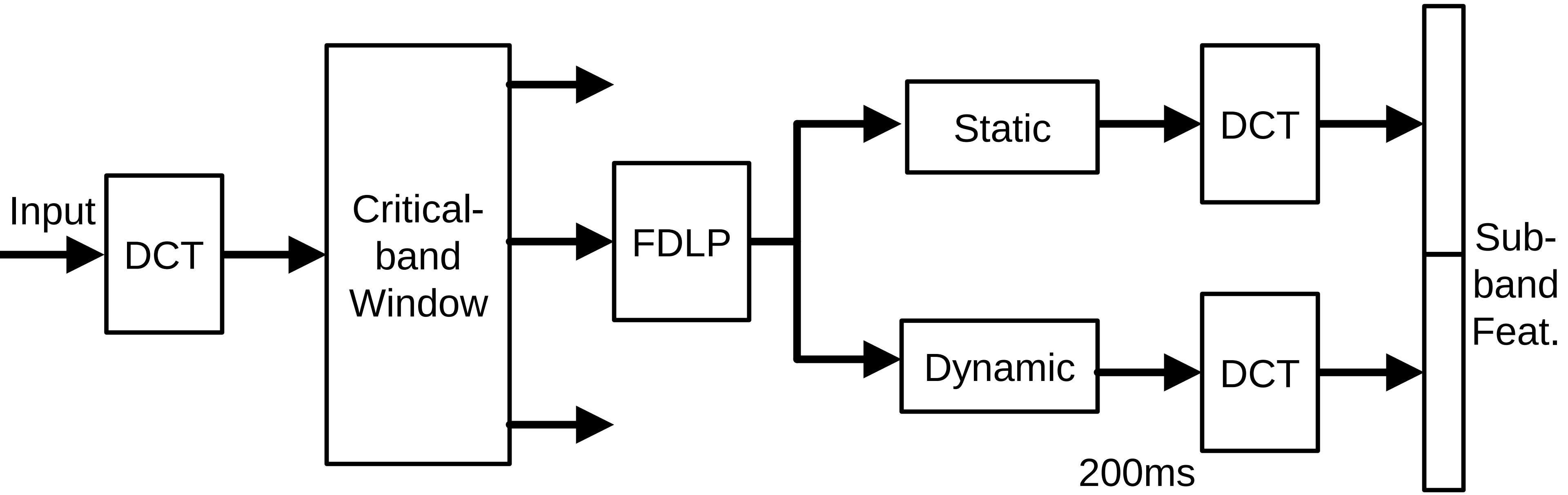


Spectrogram Comparison



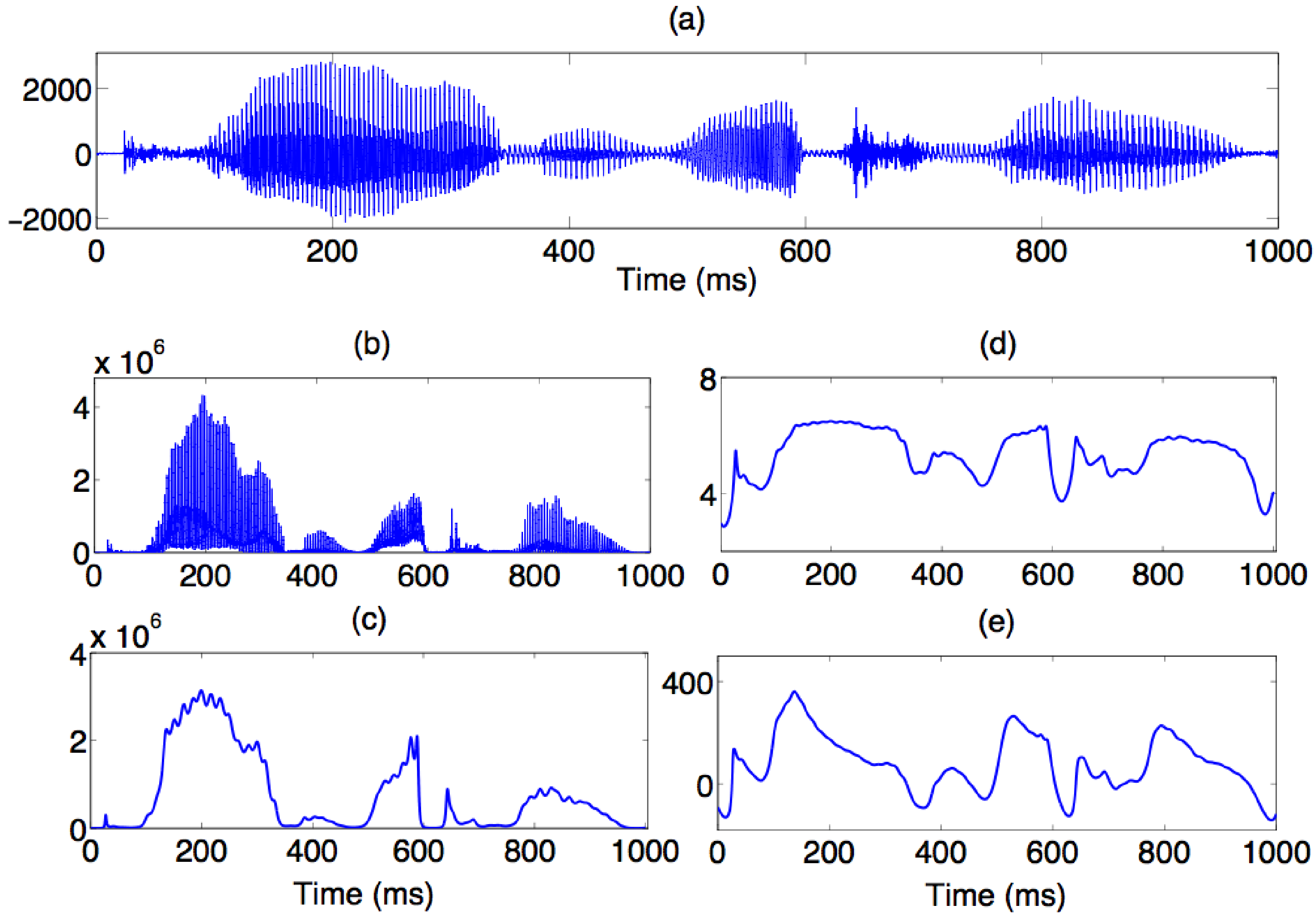
Sriram Ganapathy, Samuel Thomas and H. Hermansky, "Comparison of Modulation Frequency Features for Speech Recognition", *ICASSP, 2010.*

Modulation Feature Extraction



Modulation Features

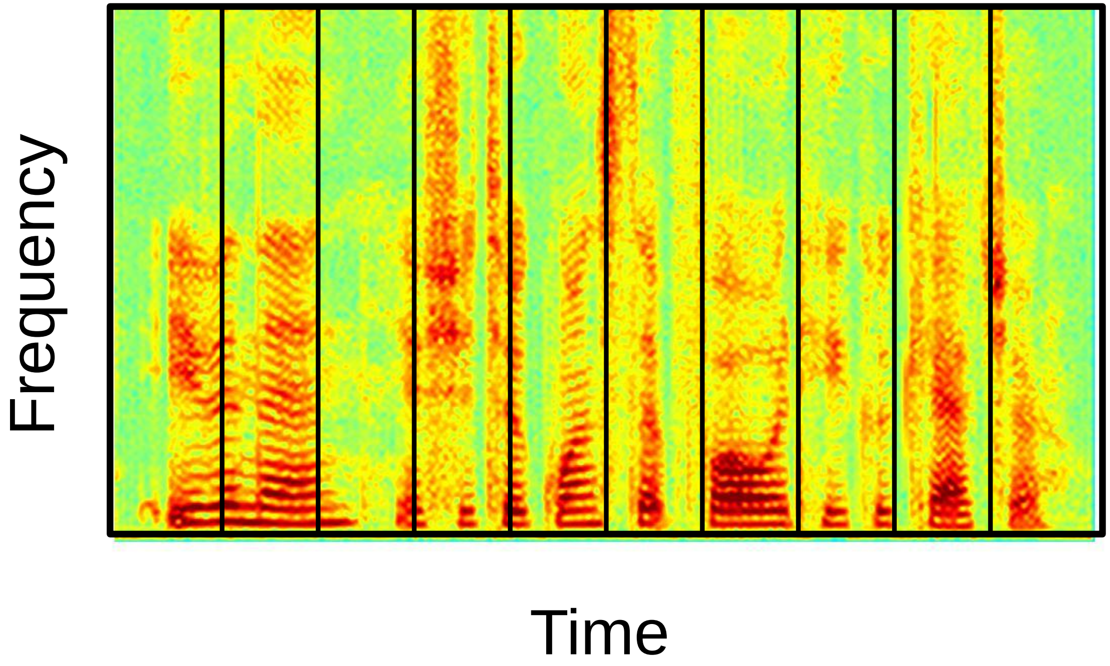
a. Signal
b. Hilb. Env.
c. FDLP Env.
d. Log comp.
e. Dyn. comp.



Sriram Ganapathy, Samuel Thomas and H. Hermansky, "Modulation Frequency Features for Phoneme Recognition in Noisy Speech", *JASA, Express Letters, 2009.*

Introduction

- Conventional signal analysis – starts with the estimation of short-term spectrum (10-40 ms).



Introduction

- Conventional signal analysis – starts with the estimation of short-term spectrum (10-40 ms).

- Spectrum is sampled at a preset rate before further modeling/processing stages.

- Contextual information is typically processed with time-series models such as HMM.

Introduction

- Conventional signal analysis – starts with the estimation of short-term spectrum (10-40 ms).

- Spectrum is sampled at a preset rate before further modeling/processing stages.

- Contextual information is typically processed with time-series models such as HMM.